A0610-1

POPQAE 48 (4) 709–850 (1984)
ISSN 0033-362X

Public Opinion Quarterly

The Public's Use and Perception of Newspapers *Leo Bogart*

The Best Seller as an Indicator of Societal Narcissism: Is There a Trend? *Lynn S. Mullins and Richard E. Kopelman*

Public Opinion du Jour: An Examination of the Spiral of Silence *Carroll J. Glynn and Jack M. McLeod*

Constituency, Party, and Representation in Congress *Benjamin I. Page, Robert Y. Shapiro, Paul W. Gronke, and Robert M. Rosenberg*

Misreporting and Vicarious Political Participation at the Local Level *Thomas J. Volgy and John E. Schwarz*

Alternative Question Forms for Attitude Scale Questions in Telephone Interviews *Peter V. Miller*

A Comparison of Methods of Increasing Parental Consent Rates in Social Research *Teresa L. Thompson*

The Content of Survey Introductions and the Provision of Informed Consent *Jeffery Sobal*

Incentives for Increasing Return Rates: Magnitude Levels, Response Bias, and Format *J. Scott Mizes, E. Louis Fleece, and Cindy Roos*

Bias in a Directory Sample for a Mail Survey of Rural Households *Frederick J. Kviz*

THE POLLS: Britain at the Polls 1945–1983 *Robert M. Worcester*

COMMENTS AND LETTERS · NEWS AND NOTES · IN MEMORIAM
BOOK REVIEWS · INDEX TO VOLUME 48

WINTER 1984

The Public Opinion Quarterly

VOLUME 48 WINTER 1984 NUMBER 4

The Public's Use and Perception of Newspapers *Leo Bogart* 709

The Best Seller as an Indicator of Societal Narcissism: Is There a Trend? *Lynn S. Mullins and Richard E. Kopelman* 720

Public Opinion du Jour: An Examination of the Spiral of Silence *Carroll J. Glynn and Jack M. McLeod* 731

Constituency, Party, and Representation in Congress *Benjamin I. Page, Robert Y. Shapiro, Paul W. Gronke, and Robert M. Rosenberg* 741

Misreporting and Vicarious Political Participation at the Local Level *Thomas J. Volgy and John E. Schwarz* 757

Alternative Question Forms for Attitude Scale Questions in Telephone Interviews *Peter V. Miller* 766

A Comparison of Methods of Increasing Parental Consent Rates in Social Research *Teresa L. Thompson* 779

The Content of Survey Introduction and the Provision of Informed Consent *Jeffery Sobal* 788

Incentives for Increasing Return Rates: Magnitude Levels, Response Bias, and Format *J. Scott Mizes, E. Louis Fleece, and Cindy Roos* 794

Bias in a Directory Sample for a Mail Survey of Rural Households *Frederick J. Kviz* 801

IN MEMORIAM: George H. Gallup, Sr., 1901–1984 *Albert H. Cantril* 807

COMMENTS AND LETTERS

On Beatty and Walter *Donald Granberg* 809

Response to Granberg *B. Oliver Walter* 811

A Note on "Knowledge Gap Effects" *Herbert Menzel* 815

Reply to Herbert Menzel *James Ettema, et al.* 816

The Open/Closed Controversy—a Postscript *Ann Pasanella* 817

NEWS AND NOTES *Philip A. Harding, Editor* 820

THE POLLS: Britain at the Polls 1945–1983 *Robert M. Worcester* 824

BOOK REVIEWS *Marc B. Glassman, Editor* 834

Steven H. Rosenstone, *Forecasting Presidential Elections* *Lee Sigelman* 834

Stephen Koss, *The Rise and Fall of the Political Press in Britain* *Leo Bogart* 835

Seymour Martin Lipset and William Schneider, *The Confidence Gap: Business, Labor and Government in the Public Mind* *Arthur H. Miller* 838

INDEX TO VOLUME 48 841

The PUBLIC OPINION QUARTERLY has been registered with the Copyright Clearance Center, Inc. Consent is given for copying of articles for personal or internal use, or for the personal or internal use of specific clients. This consent is given on the condition that the copier pay through the Center the per-copy fee stated in the code on the first page of each article for copying beyond that permitted by the U.S. Copyright Law. If no code appears on an article, the author has not given broad consent to copy and permission to copy must be obtained directly from the author. All articles published prior to 1978 may be copied for a per-copy fee of $1.00, also payable through the Center. This consent does not extend to other kinds of copying, such as for general distribution, resale, advertising and promotional purposes, or for creating new collective works.

The Public's Use and Perception of Newspapers

LEO BOGART

DAILY newspapers have been changed in many ways by the emergence of television as a major force in news dissemination and reporting. Television's vivid imagery has changed the public's perception of the personalities and places in the news, transformed the news into entertainment, and thus altered the character of press reportage. Perhaps most important, television's assertive presence has transformed news events themselves.

Michael K. Deaver, President Reagan's chief of staff and the man responsible for all White House communications operations, expressed the prevailing Washington viewpoint when he said,

> The majority of the people get their news from television, so you have to pay attention to them. You're aware of how we construct events and craft photos that are designed for 30 seconds to a minute so that it can fit into that "bite" on the evening news. We'd be crazy if we didn't think in those terms (Washington Journalism Review, 1984:25).

This philosophy affronts newspaper editors accustomed to a more searching and comprehensive style of news gathering. It also indicates the need to look at the present state of newspapers, their readership, and the

Abstract Although a widely held viewpoint is that "the majority of the people get their news from television," a national survey shows equal proportions of the population reading newspapers and watching television news on a typical weekday. The study examines the public's attitudes toward the press in general and toward the specific papers most often read. It also reports on the most commonly read elements of newspaper content and notes the changing character of local news.

Leo Bogart is Executive Vice-President and General Manager of the Newspaper Advertising Bureau. The survey reported in this article was conducted by Audits & Surveys, Inc. At the Newspaper Advertising Bureau, Albert E. Gollin directed its design and execution. Detailed reports on the survey findings are in preparation by Gollin, Charles Lehman, and Joan Lambe. Their valuable contributions are appreciatively acknowledged.

Public Opinion Quarterly Vol. 48:709–719 © by the Trustees of Columbia University
Published by Elsevier Science Publishing Co., Inc. 0033-362X/84/0048–709/$2.50

public's perception of them. This was the purpose of the study whose highlights are summarized in this paper.[1]

Dimensions of Newspaper Audiences

Newspapers are at a par with television as a source of news, if this is defined simply in terms of cumulative daily exposure. On any given day, the same proportion (67 percent) of the adult public watches at least some television news and reads at least one newspaper.[2] However, in 1971, a substantial number of people who read the newspaper on a particular day did not see any television news; today about half the public is exposed to both (Table 1).

The increased size of the audience for television news is in part a result of the fact that there is far more news on the air—for longer stretches, at more odd hours, on more stations.[3] People tune in to various types of TV programs in rough proportion to the amount of time devoted to them. News is no exception.

While the television news audience grew in the 1970s, the average weekday readership of newspapers declined (from 77 percent in 1971 to 67 percent in 1982). The problem for newspapers has been diagnosed as a decline in the frequency of readership rather than a rise in the number of nonreaders.[4] Over the course of a week, 87 percent of the public reads a newspaper on at least one day, a proportion that has not changed significantly in a dozen years. On any weekday, newspapers are read in 75 percent of all U.S. households and are in the hands of over 110 million adults. In the case of all but 5 percent, those copies are either delivered through a subscription (74 percent) or bought individually by some member of the household.

---

[1] This national probability sample survey of 1,979 adults was conducted for the Newspaper Readership Project in May 1982. The study used a multistage probability sample based on 157 primary sampling units, including two replicates. One individual aged 18 or over was interviewed in each sample household, through a random selection process. Interviewing took place Tuesday through Saturday, and interviewers carried copies of all the previous weekday's papers with circulation-to-household ratios of 5 percent or more in the county. Many of the attitude measurements were made with two versions of the questionnaire, each of which was administered to half of the total sample. A full technical description of the survey methodology is contained in a separate report, available from the author.

[2] Newspaper readership and television news exposure were based on a retrospective report on activity during each of three parts of the preceding day.

[3] In 1971, the three networks carried 29.5 hours of news per week, according to A. C. Nielsen. By 1982, this had risen to 37.5 hours. However, in the last three years the three early evening network newscasts' share of the TV audience at air time has dropped from 76 percent to 66 percent. For a discussion of television news and its appeals, cf. Bogart (1980).

[4] A more comprehensive discussion of the causes of declining readership levels, based on a similar survey conducted in 1977, may be found in Bogart (1981).

Table 1. Average Combined Daily News Exposure

	1971	*1982*
Both newspaper and TV	36%	48%
Newspaper only	33	19
TV news only	12	19
Neither source	19	14

SOURCE: Audits & Surveys, Inc. The 1971 survey data are reported in Bogart (1981):117.

There is a considerable turnover of readership from one day to the next. For every 100 readers of "yesterday's" paper, 16 will not read any paper at all "today," and 12 will be reading a different paper. However, a comparable number of "yesterday's" nonreaders will pick up a paper "today," keeping the aggregate numbers constant.

Infrequent readers are different from regular readers in certain key respects: they are younger, poorer, less well educated, more mobile, and more likely to be male or black or unmarried. They are less likely to feel that it is important to keep up with current events.

Readers' Perceptions of Newspapers and Television

The "power of the press" is not perceived as mighty, at least at the local level. We asked the public to rate nine major institutions in terms of their power to affect the well-being of people in their community (Table 2). Local newspapers and TV stations are named by only one out of five as having a "great deal" of power, half as many as those who name major corporations, organized labor, or banks.

Table 2. Power to Affect Lives of People

Here is a list of several kinds of organizations or groups. Please tell me how much power you think each one has to affect the lives of the people living in (insert city). Would you say that —— [order rotated] has a great deal, some, hardly any, or no power to affect the well-being of the people?

% Saying "Great Deal of Power"	
Federal government	68%
State government	55
City government	44
Major corporations	43
Organized labor	43
Banks	41
Local newspapers	21
Local churches	19
Local TV stations	18

Curiously enough, only one reader in four acknowledges that his or her newspaper "prints too much bad news"; infrequent readers are only slightly more likely to say so, though other studies suggest that many of them are turned off by the complexities and unpleasant aspects of news itself (Bogart, 1981:61–63).

Where do people get their information about the issues that concern them? When it came to the problem they identified as the most important one facing the country (the economy, in the eyes of two out of three), television was named as the information source that is relied on most by 45 percent, newspapers by 30 percent, radio by 6 percent. But newspapers led TV among college graduates by 42 percent to 30 percent.

We also asked people where they *first* found out about whatever they considered the biggest current news story. (Half of them named the war in the Falklands.) Whatever news story was mentioned, television was named by 51 percent, newspapers by 27 percent, and radio by 17 percent as their first news source on that story.

The realities of information flow may be quite different from how the public perceives them. Television has an edge over newspapers when it comes to specific, high-impact stories. Four out of five people agree that "seeing the news on TV makes it easier to understand what's happening." But a clear majority, 53 percent, say they do not get enough news from TV and want the added details from newspapers on the big stories, while 30 percent say they get enough from the TV news.

One aspect of TV's appeal, as reflected in these findings, is that people feel closer to the television channel they watch most often than to the newspaper they read most often. As Table 3 shows, TV has gained significantly on this dimension in the past 20 years. What accounts for this change? The channels watched most often by most people are network affiliates, with only a modest amount of locally originated programming. It is the local news shows that help give a television channel its identity. Thus, the answers to this question show how powerfully TV newscasters affect overall ratings of media performance.

Can newspapers compete with television news in this area by building public recognition of their own by-lined writers and columnists? Forty-

Table 3. To Which Do You Feel Closest?

Generally speaking, which do you feel closest to—the newspaper you read most often, the radio station you listen to most often, or the television channel you watch most often?

	1961	*1982*
Newspaper	40%	28%
Television channel	33	40
Radio station	14	15

one percent of the readers say they regularly follow a particular columnist or reporter. Of these, 72 percent feel they know what he or she is like as a person, and 76 percent say they would like that writer as a friend. Some newspaper managements encourage reporters and editors to become public personalities. Others discourage the development of "stars" on the grounds that it leads to salary demands, inflated egos, and defections to other employers. The evidence seems to indicate that the reader's sense of the character of the writers strengthens the attachment to a paper.

But how strong are the links between the public and the news media? About one out of three say they would "feel lost" if they couldn't buy a newspaper or if they couldn't watch the news on TV. And one in five say they could "easily do without" the newspaper and without television news. The individuals who give this response are to a large extent the ones who are not exposed to that medium on an average day. But of the remainder, nearly half think that while they might miss newspapers or television news, they could get along without it. This suggests a casual or ambivalent attitude toward the news itself that, among other things, erodes the regularity of newspaper reading.

Ratings of Press Performance

Attitudes toward the press in general are less meaningful than the ratings people give to specific papers in their own communities—the ones they read "yesterday," or would be most likely to read. They were provided with a representative range of topics—certainly not a complete one—and asked to rate their own paper's coverage of each. As Table 4 shows, there is considerable variation in the percentages who think that their paper does a "very good" job in each of these areas.

There is no news topic about which a majority say their newspaper does a "very good" job of providing information. But ratings of coverage as "not good at all" are rare. Metropolitan papers get higher ratings for their coverage of sports, crimes, and economic news; smaller papers naturally get a higher rating on neighborhood news.

The local newspaper's coverage of different kinds of news represents only one dimension of performance. Respondents were also asked to rate their own paper nine ways, using a scale on which 1 was "extremely poor" and 9 was "excellent." (Table 5). Rated lowest of all was the balance in coverage, that is, "how fair or impartial it is to all sides."

Who are the people who rate their local newspaper *below* average on fairness? It might be expected that they would be concentrated among those who are least familiar with it, who read it least. However, the critics are mainly concentrated among the better educated, who are, in fact, the

Table 4. Evaluation of Main Newspaper

Using the phrases on this card, please tell me—for each of the topics I read to you—how good a job the newspaper you read most often does in helping to keep you informed. Let's start with ____. Would you say the job done by that newspaper is very good, fairly good, not too good, or not good at all in helping to keep you informed about ___?

% Who Say "Does a Very Good Job" on:	
Crimes	44%
Sports	42
Inflation	39
Unemployment	35
Other countries	30
Local government	29
President Reagan	25
Young people & drugs	24
Neighborhood news	18
Personal health	17
Pollution problems	14

Table 5. Average Ratings of Main Paper (Scale of 1–9)

Here's something a little bit different. I'd like to have you rate the daily paper you read most often, using this scale. As you see, it goes from one to nine, with "1" at the bottom or extremely poor end, and "9" at the top or excellent end of the scale. How would you rate that newspaper in terms of ___?

How hard it is to read and understand	7.41
How good it looks	7.06
The quality of the people who work for it	7.05
The kind of value it gives its readers for the money	7.05
How good a job it does in bringing to light the things the public ought to know about	6.96
How interestingly it is written	6.89
How accurately it presents the news	6.82
How genuinely interested the paper is in this community	6.52
How fair or impartial it is to all sides	6.13

most faithful readers and who make the heaviest demands for high-quality performance.

If fairness is newspapers' weakest point, does the public think they keep editorial opinions separate from their news coverage? Only 43 percent of the readers give them an unqualified clean bill of health on this score, while 32 percent think that editorial opinions affect news coverage at least sometimes. Among newspaper readers, 78 percent say they read editorials; 35 percent try to do so regularly. (Readers of editorials are somewhat more likely to feel that opinions are kept separate from news coverage.) Most readers, whether frequent or infrequent, agree that editorial or op-ed pages offer a wide range of opinions. But 25 percent say they "mostly reflect just one basic viewpoint."

How do readers describe the political leanings of their own paper? Across the country, 40 percent describe it as middle of the road. Of the remainder, the balance tilts 24:16 percent toward describing it as conservative rather than liberal. Newspaper readers in big towns correctly perceive their papers to be more liberal than rural readers do. As for the television news program that each respondent watches most often, more (26:19 percent) regard it as liberal than as conservative.

The way people describe their own newspaper's politics is not related to their personal sympathy for its editorial viewpoint. Overall, almost two out of three say they agree with that viewpoint, whether they consider the paper to be mostly liberal, mostly conservative, or middle of the road.

Content and Readership

Occasionally it is suggested that readers are turned off by the size of the paper, by their inability to cope with the volume of printed matter. (The size of the typical paper ranges from 15 pages for those with under 10,000 circulation to 100 pages for those with over 500,000 circulation.) Regardless of the size of the town they live in, three out of four readers like the paper the way it is. Of the remainder, overall, more would prefer it thicker than thinner.

News continues to be the main attraction. Faced with an imaginary choice between a paper that was mostly news or mostly features (with just a summary of the news), 49 percent opt for a paper that is mostly news; 20 percent for a paper that is mostly features, and 26 percent volunteer the answer that they want both—which is exactly what editors are giving them. That same point is brought out when people are asked whether they look forward more to the regular features or to the news items. Fifty-nine percent choose the news items; 40 percent the regular features.

Daily newspaper circulation has been static in recent years, while the feature-filled Sunday papers have kept on growing at a healthy rate. People were asked which they would prefer to keep if they had to make a choice. Only a minority (29 percent) was willing to sacrifice the daily paper for Sunday, but that minority gets up above one-third among people under age 30, and it remains around one-third among people age 30 to 44.[5]

The proportion of the newshole devoted to different categories of content, based on a count of items, has not changed substantially since

---

[5] Previous research has shown that the Sunday paper's appeal to the younger reader is based on its depth of special interest material (Bogart, 1981:126–27).

1977, or for that matter since 1971. And the relative percentages of readers who report reading typical editorial items in different categories of content have stayed much the same since 1971.

Analysis of newspaper content shows that much more space is devoted to sports or business news than to news of home or fashions (Bogart, 1981:159). Does this mean that women feel the paper favors men's interests? In fact, only 10 percent have this opinion (compared with 17 percent for men); 5 percent think there is more in the paper for women than for men, but 81 percent feel there is about the same for both.

Three out of five people say that the paper offers the same to the better and to the less educated. Among the minority, far more say the better educated are favored. But people who have not completed high school are likely to say that there is the same amount in the paper for them as for the better educated. In other words, they tolerate a higher standard.

Page Opening and Item Readership

General descriptions of reader attitudes and preferences are less revealing than reports of actual reading activity. Following a procedure used in a series of similar studies, readers were taken through 10 pages (four randomly selected spreads, plus the front and last pages) in the papers they read "yesterday" (leaving aside spreads devoted entirely to classified advertising). They were asked if they had opened and scanned these pages in their original reading and were further asked which items on those pages they had read. (Items of less than five inches were covered only in a subsample of interviews.) Those who had not read a paper "yesterday" were asked about their *interest* in items on the sampled pages of the paper they usually read.

Seventy-seven percent of all sampled pages were opened, a proportion similar to that found in previous national surveys in 1963, 1971, and 1977. This percentage does not vary substantially by sex, age, income, or education. Pages that carry every kind of content in the paper are opened by the vast majority of readers.

For the average paper, readers reported reading more than one-third (35 percent) of all the news and editorial items over five inches on opened pages. For some categories of editorial content the typical item was read by nearly half: items such as news briefs or summaries, crime, war/defense. Women gave personal advice columns the highest readership figure (57 percent) achieved by any category of content for either sex. Different kinds of readers respond to different kinds of content, but every kind of content finds at least some readers in every population segment.

Among readers, attention levels are much higher for illustrations (photos) or for stories accompanied by illustrations than they are for straight text. Cartoons, comics, and graphics fall in between. Among nonreaders, the interest pattern is similar. Reading and interest scores are both about the same for editorial items on right and left pages. Stories arouse slightly more reader attention and interest when they appear on pages that consist over half of advertising than on pages on which text predominates.

Front-page stories get much better readership than the average, as well as higher interest among nonreaders. Stories on the back pages of the paper or of separate sections get slightly more reading or interest than does the average item. But apart from that, reading levels vary remarkably little as we move from the front to the back of the paper.

Similarly, readership and interest are appreciably higher for longer articles (many of which include illustrations), but items of different sizes below 18 column inches show virtually no difference. This supports a conclusion from earlier studies that people going through the paper engage in a systematic search for what is relevant to them and will read it or judge it as interesting or not, regardless of size.

Near and Far Horizons

The preeminent importance of local news has long been accepted by American editors. For a daily of 8,000 circulation in a rural area, local news *is* community news; it deals with personalities, people, and situations to which almost any reader can relate directly and fit into the mosaic of common everyday experience.

In a huge metropolitan region, events that make news are those that meet general criteria of interest and importance. Apart from public figures, the protagonists of stories are not known personally to the readers.

A majority of the American public reads metropolitan papers, but a majority of the papers published falls somewhere between the extremes of size. Their editors must therefore distinguish between what is inherently newsworthy for most of their readers (either because of its areawide impact or dramatic character) and what hits home to those readers who live in the specific neighborhood where the event takes place or who share some common attribute with those involved.

The distinction between what is local and what is not is increasingly hard to make. Readers respond to stories in terms of their relevance, importance, or human interest, regardless of their origin. They expect newspapers to cover the world as well as their own localities. In fact, a majority of the general public and three-fourths of the college graduates say they are more interested in international and national news than in the news of their own city or town.[6] And within metropolitan areas, interest in local news is less among suburban readers than it is among central city residents.

When asked about 25 different categories of content, 84 percent of the readers say they usually read international or world news, only slightly below those mentioning news about jobs, inflation, and the economy (86 percent) and news about the local community (89 percent).[7] When asked what they care most about from among categories of news usually read, 36 percent pick international or world news, 31 percent news about the local community. The high level of interest in international news cuts across every segment of society.[8]

Compared with the typical item of state or local news, national news stories score 11 percent higher in reported readership and international news stories 29 percent higher.[9] No matter where they appear in the paper, the readership of world news is markedly ahead of local stories. And the advantage holds for stories of every size, though international stories tend to run longer. (Among people who had *not* read yesterday's paper, interest in the typical international story also runs 29 percent ahead of interest in local news items.)

Of all the editorial items five column inches or larger, 14 percent were of local and state news and so were 17 percent of the smaller items. By contrast, national/international news items account for 11 percent of all the items over five inches and for 10 percent of smaller items.

Of all the stories classified as local news, 23 percent deal with the actions of local governmental authorities, 22 percent are crime stories, 21 percent are school news, 11 percent deal with accidents, and 7 percent are traditional "society news"—births, marriages, engagements, "chicken dinners," etc.

---

[6] In 1961, a majority of the public were more interested in the local news.

[7] Four of the 25 categories readers were asked about are advertising rather than editorial content: "supermarket ads," "store ads for clothing," "classified ads or want ads," and "movie ads." The results indicate that advertising is actually better read than some editorial content. Over six out of ten readers (64 percent) imply that they usually read supermarket ads, as compared with 47 percent reading political opinion columns and 29 percent reading book reviews.

[8] It dips only slightly among people 18 to 24, those with low income and education, and those living in rural communities. Not surprisingly, local news is of greatest interest in rural areas and of least interest in large metropolitan areas. It is of most concern to the people who have lived longest in a community. But it is not the exact obverse of interest in the world scene. Interest in local news is highest among people in their thirties and forties. It is higher among women than among men. It is higher among people in the middle income range than among those at either extreme. It is not related to the level of education, but it is somewhat greater among frequent newspaper readers. The biggest papers carry 27 percent more national/international than local news, small papers 28 percent less, but hard news (international, national, and state/local items) accounts for only 21 percent of all the items in the small papers and 32 percent in the big ones.

[9] In the 1977 study, which used interest rather than recognition to measure individual items (obviously for a different mix of stories), national and international items scored about 50 percent higher than local news items.

When we break local news into its components, these draw quite a range of responses from the public. And the *intensity* of appeal of these locally originating types of coverage varies substantially among those who scan them regularly.

Conclusions

The survey reported in this paper catches a passing moment in a communications environment that is undergoing continuing and rapid change. The evidence documents television's growing importance but does not support the assumption that it has become the preeminent source of news. The issue of preeminence is itself meaningless in a society where half the public is exposed daily both to newspapers and to television news. Newspapers are part of the life of nearly nine out of ten Americans and touch two out of three on a typical day. And even in the era of "sectionalization," when it is easier in theory for people to set aside content they want to avoid, nearly four out of five readers report looking at any given page. In the perceptions of a majority of the audience, the two media complement each other, and newspapers' ability to cover the news of the area in detail and in depth remains a major advantage.

Many newspapers have tried to compete with television through an increased emphasis on entertaining features. But it is the *mix* of diversions with the realities, often the unpleasant realities, of the news that has maintained the reading habit for centuries and preserved newspapers' place in the shaping of public opinion. A small-town publisher once asked the Chicago *Tribune's* Joseph Medill the secret of success in newspapering. "Publish the news," said Medill. That still seems like good advice.

References

Bogart, Leo
1980 "Television news as entertainment." In Percy Tannenbaum (ed.), The Entertainment Functions of Television. Hillsdale, NJ: Lawrence Erlbaum Associates.
1981 Press and Public: Who Reads What When Where and Why in American Newspapers. Hillsdale, NJ: Lawrence Erlbaum Associates.
Washington Journalism Review
1984 "Michael Deaver rates the president's press." Washington Journalism Review 6 (April): 23–25.

Public Opinion du Jour: An Examination of the Spiral of Silence

CARROLL J. GLYNN AND JACK M. McLEOD

"PUBLIC OPINION," as a concept and as a process, has been considered a key factor in the outcome of political and social issues for centuries. Theorists such as Allport (1937) and Almond (1950) regarded public opinion as a transitory phenomenon, but recent evidence indicates that it may not be as volatile or transitory as some critics claim, and that when changes do occur, they are not random or capricious (Page and Shapiro, 1982). However, during short-term events such as elections, there may be substantial changes in opinion because people are continually receiving new information about political candidates, information that helps them adjust their voting decisions.

It is in this context of high public uncertainty combined with an increase in the flow of communication that Noelle-Neumann's "spiral of silence" research (1977) is important. The theory of the spiral of silence states that one's perception of the distribution of public opinion motivates one's willingness to express political opinions. This act of self-expression then changes the "global environment of opinion, altering the perceptions of other persons and, ultimately, affecting their willingness to express their own opinions" (Taylor, 1982). Individuals who notice that their own personal opinions are spreading will voice these opinions self-confidently in public; those who notice their opinions are "losing ground" will be

Abstract This study investigates the spiral of silence theory using perceived vote outcomes as well as actual vote behaviors. The first major hypothesis—that those who see their position as gaining support will be more likely to discuss while those who see their position as losing support will be less likely to discuss—was supported by the data. Some support was also found for the hypothesis that individuals perceiving support for a certain candidate would be more likely to express a preference for the candidate. Results also seemed to justify the Noelle-Neumann distinction between the resistant hardcore and other types of citizens. The relative independence of perception of opinion judgments from own preferences also was investigated and support was found for this difference.

Carroll J. Glynn is an Assistant Professor in the Department of Communication Arts, Cornell University. Jack M. McLeod is Baier-Bascom Professor in the School of Journalism and Mass Communication, University of Wisconsin.

Public Opinion Quarterly Vol. 48:731–740 © by the Trustees of Columbia University
Published by Elsevier Science Publishing Co., Inc. 0033-362X/84/0048–731/$2.50

inclined to adopt a more reserved attitude and remain silent. Noelle-Neumann labels this process the "spiral of silence." The basic premise of Noelle-Neumann's work is that through social interaction, people influence each other's willingness to express opinions. That is, she believes that to the individual "not isolating himself is more important than his own judgment" (Noelle-Neumann, 1977).

Taylor (1982) states that the four variables in Noelle-Neumann's research are: "one's opinion on an issue; one's perception of the predominant public opinion; one's assessment of the likely future course of public opinion; and one's willingness to support one's opinion with action, verbal statements, or other signs of commitment." One's opinion and one's assessment of the predominant public opinion are both assumed to influence one's judgment of the future course of opinion. Willingness to express one's opinion depends on an interaction between one's opinion (majority or minority) and one's perception of the future trend in public opinion.

We investigated Noelle-Neumann's spiral of silence theory in the present study, using perceived vote outcomes as well as actual vote behaviors. Noelle-Neumann states that the tendency for some individuals to speak up and for others to remain silent starts a "spiraling process" which increasingly establishes one opinion as the prevailing one. In addition, she notes that in relation to election campaigns, this spiraling process of opinion expression (measured with the question "Who do you think will win the election?") can lead, in the final phase of the campaign, to an increase in the *number of supporters* of a political party and "thus the decisive margin for the election victory" can become obvious (Noelle-Neumann, 1977:148).

Improved measures (described below) from the 1980 presidential election campaign were used to operationalize some of Noelle-Neumann's major variables in order to study, more thoroughly, some of the basic premises of the spiral of silence. Two major hypotheses were tested and the following predictions made:

> H1: Those individuals who believe there is a trend in support of their position are more likely to express their opinions publicly than are those who do not.
>
> H2: The more individuals see a candidate as winning an election, the more likely they will be to express preference for that candidate.

In addition, the present study attempted to investigate a briefly mentioned distinction in the Noelle-Neumann spiral of silence theory—the "hardcore" opinion holders. Noelle-Neumann (1974:48) states that the hardcore are "not prepared to conform, to change their opinions, or even to be silent in the face of public opinion." Yet, Noelle-Neumann does not employ this concept when analyzing her data. In the present study, the hardcore variable is investigated as a control variable, as a main effect, and as an interaction term. In this way it is possible to analyze Noelle-Neumann's assumption that hardcore individuals are not influenced by the prevailing "climate of public opinion" or by the "spiral of silence" in the same way as the majority of individuals.

Study Design

Telephone interviews were conducted with a panel of 98 eligible voters in Dane County, Wisconsin at three time points—two prior to and one after the 1980 presidential election. The first interview was taken before the initial September debate, the second in late October, and the third in the week after the November election. The second and third waves also included a replenishment sample of 367 additional respondents not previously interviewed. Random digit dialing procedures were used with proportionate probability sampling stratified according to the number of private lines at each exchange. Respondents were screened at each selected number to ensure their eligibility to vote.

Two types of perceptions of public opinion were used to form the critical independent variables for the study. In both the September and October interview waves, respondents were asked to estimate what percentage of the vote each of the three major candidates would receive in November (Estimated Percent). They were also asked whether each candidate was gaining support, staying the same, or losing support (Support Change) in both the September and October waves. The data thus included separate estimated percentage and support change measures for each of the three candidates at two time points.

Separate measures of support change were combined into a single index of Support Discrepancy. Each respondent's presidential choice (Carter, Reagan, or Anderson) was compared to that person's judgment of October support change for the preferred candidate. High discrepancy (3) was assigned cases where the preferred candidate was seen as losing support, moderate discrepancy (2) was given where that candidate was judged as staying the same, and low discrepancy (1) was attributed where the chosen candidate was perceived as gaining support.

Noelle-Neumann's conceptual distinction between the hardcore opinion holders and the rest of society was also investigated despite a lack of operationalization in her previous research. The variable was indexed here by using the degree of vote *consistency* over time. For the larger panel ($N = 465$), those who said in October that they were certain to vote for a given candidate and who did in fact vote for him in November were called Hardcore, while those less certain or who changed were combined into a Combined Leaner and Nonhardcore category. A slightly more sensitive index was used with the small panel ($N = 98$), for whom we had September vote preference measures. Those who unequivocally preferred a given candidate in both September and October and who reported voting for that same candidate in November were classified as Hardcore. Those who were leaning toward a candidate in September and/or October and who voted for that candidate in November were called Leaners, and those showing greater change, switching candidates between rounds, for example, were labeled Nonhardcore.

At first glance there would appear to be a problem with conceptualizing Hardcore in this fashion; aspects of voting behavior enter into both the contingent Hardcore variable and the dependent Vote Preference variable. However, the confounding is not real. The measurement of Nonhardcore and Leaners reflects the degree of *change* or lability of vote preference over time regardless of direction, whereas vote preference and actual vote reflect *direction* and *intensity* of the intention at a given point in time.

In order to test Noelle-Neumann's first major hypothesis of the spiral of silence, a special set of questions was devised based on hypothetical social situations comprised of other people, potentially congruent or noncongruent with the respondent's vote preference, discussing these preferences. We asked:

> If you were at a social gathering with people you know, and they were discussing the presidential campaign, how likely would you be to enter into the conversation if they were Reagan supporters? Would it be very likely, somewhat likely, not very likely, or not at all likely that you would join the conversation?

The question was repeated for the situation where the social gathering consisted of Carter supporters. For respondents supporting either Reagan or Carter in October, the Reagan and Carter social gathering questions were identified as either congruent or noncongruent according to the vote preference (or nonpreference) of the respondent. For supporters of the two leading candidates, then, the dependent variable could be assessed by comparing the likelihood of entering the discussion in the two types of social gatherings. Time limitations prevented the use of this question for a gathering of Anderson supporters. Anderson backers were thus excluded from this analysis because a congruent situation could not be analyzed for them. Sample size limitations led us to use the larger panel only for this analysis.

The major dependent variable, Vote Preference, was measured in September and October by creating an index for each of the three major candidates using a five-point scale: sure of voting for the given candidate (5), leaning toward the candidate (4), undecided (3), leaning toward another candidate (2), and sure of voting for another candidate (1). In the postelection survey, voting was measured by reports of having either voted for or not voted for a given candidate.

Hierarchical least-squares regression was the basic method used in the analysis of vote preferences. The control variable, party affiliation, was introduced as the first predictor. The main effects of the two perception of opinion variables (Estimated Percent and Support Change for the given candidate) were then entered as a second predictor block. The third block was Hardcoreness as a main effect. Finally, a fourth block comprised of the two interactions of Hardcoreness with the perception of opinion variables was introduced. Significance tests (*F* values) were used for individual Betas and for blocks of predictor variables.

Two relatively stable political variables, Party Affiliation and Political Interest, served primarily as control variables. Party affiliation was indexed as a seven-point scale as follows: strong affiliation with the Republican party, 7; Moderate (not strong) Republican, 6; Independent Republican, 5; Independents without a leaning, 4; Independent Democrats 3, and so forth. Political interest was measured by asking: "Generally, would you say you are very interested in politics, interested in politics, somewhat interested in politics, or not at all interested in politics?"

In order to examine various communication sources for their potential influence on perceptions of public opinion, respondents were asked: "We're interested in which source you use most for political information. Would you say you use newspapers, friends or family, television, magazines or some other source most for current events and political information?" After choosing which they used most, the respondents then ranked the other sources. The rankings provided data for three indices: Television Reliance, Newspaper Reliance, and Interpersonal Reliance.

Results

PUBLIC EXPRESSION OF OPINION

The first hypothesis was tested by examining the effects of support discrepancy on the likelihood of entering into a political discussion with people with similar (congruent) or dissimilar (noncongruent) viewpoints in hypothetical social situations (Table 1). As noted previously, support discrepancy was measured on a three-point scale.[1]

---

[1] Noelle-Neumann traditionally has used a hypothetical situation where respondents assume they are sitting next to a stranger in a train. However, because trains are not a primary mode of transportation or a common method for meeting individuals in the United States, the authors felt it was preferable to use a hypothetical *social* gathering for the present study. In this fashion, respondents could be placed in a situation where they are faced with individuals with similar opinions and in situations with individuals with opposite opinions.

Table 1. Likelihood of Entering into a Campaign Discussion at a Congruent and Noncongruent Social Gathering, by Level of Support Discrepancy Among Hardcore, Combined Leaner and Nonhardcore Group and All Panel Respondents

		Social Gathering					
		Congruent		*Noncongruent*			
	Support Discrepancy	*Mean*	*S.D.*	*Mean*	*S.D.*	*Diff.*	*N*
Hardcore	High	3.36	(1.57)	3.00	(1.59)	+.36	(28)
	Medium	3.81	(1.16)	3.45	(1.39)	+.36	(86)
	Low	3.93	(1.27)	3.60	(1.34)	+.33	(95)
		3.80	(1.28)	3.46	(1.36)	+.34	(209)
Combined leaner and nonhardcore	High	3.74	(1.29)	3.26	(1.24)	+.48	(19)
	Medium	3.43	(1.34)	3.05	(1.34)	+.38	(40)
	Low	3.85	(1.06)	3.47	(1.26)	+.38	(68)
		3.70	(1.15)	3.31	(1.29)	+.39	(127)
All panel respondents	High	3.51	(1.42)	3.11	(1.43)	+.40	(47)
	Medium	3.69	(1.18)	3.33	(1.37)	+.36	(126)
	Low	3.90	(1.17)	3.55	(1.30)	+.35	(163)
		3.76	(1.23)	3.40	(1.33)	+.36	(336)

NOTE: High Support Discrepancy are those seeing their preferred candidate as losing support. Medium Support Discrepancy are those seeing their candidate as staying the same, while those seeing a gain in support are categorized as Low Support Discrepancy.

As expected, respondents ($N = 336$) were more likely to say they would enter into a congruent political discussion than into a noncongruent one (3.76 vs. 3.40 for all panel respondents, $F = 16.60, p = .001$). Those with lower levels of support discrepancy were also more likely than others to want to express their point of view (3.72 vs. 3.51 vs. 3.31, $F = 4.58, p < .05$). The Hardcore did not differ from other respondents in this respect (3.63 vs. 3.50, $F = 1.74$, n.s.) and no combination of these three variables produced a significant interaction.

PERCEPTUAL INFLUENCES ON VOTING

In order to test the second hypothesis, that individuals who see a candidate as winning an election would be more likely to express a preference for that candidate, regression analyses were conducted. As shown in Table 2, vote preferences expressed in September and in October and actual vote in November were regressed on party affiliation, the perceptions of opinion variables, the Hardcore variable and the interaction between the perception of opinion variables and the Hardcore variable.

As expected, party affiliation accounted for a large and significant amount of the variance at all time points when predicting to Carter vote preference and Reagan vote preference, and a lesser amount when predicting to the September Anderson vote preference. The Democrats were more likely to prefer Anderson than were Republicans.

Table 2. **September and October Vote Preferences (Perceived Outcome) and November Vote Regressed on Party Affiliation, Estimates of Candidates' Vote Percentages, Support Change, Hardcoreness, and Hardcoreness-Percentage/Support Interactions ($N = 78$)**

	Carter		*Reagan*		*Anderson*	
	Simple r	*Incr. R^2%*	*Simple r*	*Incr. R^2%*	*Simple r*	*Incr. R^2%*
September preference						
Party affiliation	−.34*	11.4*	.64*	40.8*	−.33*	11.0*
Estimated percent: Sept.	.37*	5.7*	.48*	7.9*	.05	0.0
Support change: Sept.	.15	0.8	.27*	0.4	.13	1.5
Hardcoreness	.07	0.9	.14	0.9	−.26*	5.7*
H × estimated percent	.13	5.0*	.11	0.0	−.14	0.0
H × support change	.22*	6.3*	.12	0.3	−.23*	0.4
Variance accounted for by:						
Percent/support var. (4)		17.8*		8.5*		1.9
All variables (6)		30.1*		41.3*		18.6*
October preference						
Party affiliation	−.55*	30.6*	.63*	40.3*	−.16	2.6
Estimated percent: Oct.	.51*	12.0*	.30*	1.1	−.04	0.0
Support change: Oct.	.10	1.1	−.01	2.0	.22*	4.7*
Hardcoreness	−.12	0.2	.21	0.3	−.08	0.3
H × estimated percent	−.21	0.2	.13	0.2	−.08	2.3
H × support change	.20	1.7	.17	0.0	−.22*	0.6
Variance accounted for by:						
Percent/support var. (4)		16.0*		3.3		7.6
All variables (6)		46.8*		43.9*		10.5
November vote						
Party affiliation	−.59*	35.2*	.68	47.1*	−.11	1.1
Estimated percent: Oct.	.36*	4.4*	.31*	1.5	−.14	1.7
Support change: Oct.	.07	0.6	.09	0.2	.13	2.0
Hardcoreness	−.23*	0.7	.18	0.0	.18	3.9
H × estimated percent	−.28*	1.5	.09	1.8	.20	0.3
H × support change	−.24*	0.6	.10	0.1	.07	0.2
Variance accounted for by:						
Percent/support var. (4)		7.1		3.6		4.2
All variables (6)		42.0*		50.7*		9.2

* $p < .05$

In the case of Carter, one of the perception of opinion variables (Estimated Percent) accounted for a significant amount of the variance at all three time points. That is, the more respondents saw Carter as having a larger percentage of the vote, the more they preferred him. Both perception of opinion variables accounted for a significant proportion of the variance when predicting to Reagan vote preference, but only in September (8.5 percent). These variables did not account for much of the variance when predicting to Anderson vote, except in October (7.6 percent), when those seeing him as holding even or gaining were more likely to prefer him.

The effect of the interaction of the Hardcore variable with the two perception of opinion variables accounted for a significant proportion of the variance in only one of the six comparisons in Table 2. In September, Hardcore and perceptions of public opinion combined to produce a strong vote preference for Carter. The two interactions accounted for 11.3 percent of the Carter vote preference variance.

The evidence, overall, tends to uphold the hypothesized connection between perceived public support for a candidate and vote preference for that candidate.

Discussion

It is obvious from the results of this study that the nature and development of public opinion is a complex cognitive and societal process. The first major hypothesis—that those who see their position as gaining support will be more likely to discuss while those who see their position as losing support will be less likely to discuss—was supported by the data. Some support was also found for the second major hypotheses—that individuals perceiving support for a certain candidate would be more likely to express a preference for that candidate. In the present study these perceptual variables were found to influence not only vote preferences for Carter in September and October, but also reports of actual vote for him in November. They also predicted to vote preferences for Reagan in September and for Anderson in October. The interaction of Hardcoreness with the two perception of public opinion variables significantly affected vote preferences for Carter but not for Reagan or Anderson. While this is somewhat surprising (we would have expected perceptual variables to influence Anderson votes), perhaps for the relatively well-informed liberal voters in this sample it was a matter of estimating whether Carter had a chance because they knew Anderson did not.

Although such substantial connections between perceptions of public opinion and personal vote preferences have been uncovered, it is important to be wary in asserting the direction of causation. Vote preference may lead to holding distorted estimates of public opinion, although panel analyses in previous research make a case for treating vote preference as a dependent variable (Glynn and McLeod, 1982). The lack of relationship between party affiliation and perceptions of opinions also argues for this casual direction.

The results seem to justify the Noelle-Neumann distinction between the resistant Hardcore and other types of citizens. Interactions of Hardcoreness with other variables were frequent enough to argue for this distinction. In addition, other analyses not shown here have found the Hardcore to be different from other respondents in their higher level of campaign discussion, in their reports of lesser influence of this discussion on their voting decisions, and in their attribution of more importance to candidates' issue stands and personal qualities. They also were less likely to see the candidates' chances of winning as a factor in their choice, providing some validation to our findings reported here.

Most of the variance in voting preferences was not accounted for by the perception of public opinion variables. That is, the perceptions of public opinion seem to operate with relative independence from the individual's own position. To illustrate this further we have compared the aggregation of individual opinion (Own Preference) to five other aggregated expressions of opinion: individual perceptions of current opinion (Perceived Opinion); perceptions of vote outcome (Perceived Outcome); current national Gallup poll data (Poll Data); panel respondents' November vote (Own Vote); and national vote in November (Actual Vote) (Table 4). Table 4 does provide support for the separation of perceptions of opinion from aggregated individual opinion (usually known as "public opinion"). In September, for example, while the aggregate of individual responses gave Carter a clear lead over Reagan (41 percent vs. 32 percent) with Anderson a close third (27 percent), respondents perceived the current opinion (Perceived Opinion) as being close between Carter (42 percent) and Reagan (40 percent), and placed Anderson a more distant third (18 percent). All three estimates were within 3 percent of the most recent September Gallup poll prior to the first interview wave, considerably closer than to the aggregate of the respondents' own preferences.

The relative independence of perception of opinion judgments from own preferences also is seen in the results for the respondents' percep-

Table 3. Vote Preferences, Perceptions of Current Opinion, Perceptions of November Outcome and Gallup Poll Data in September and October, and Panel Vote and National Vote in November ($N = 90$)

	September				*October*			*November*	
Candidate	*Own Pref.*	*Perceiv. Opinion*	*Perceiv. Outcome*	*Poll Data*	*Own Pref.*	*Perceiv. Outcome*	*Poll Data*	*Own Vote*	*Actual Vote*
Carter	41%	42%	46%	41%	46%	45%	46%	45%	42%
Reagan	32	40	43	43	36	45	45	41	51
Anderson	27	18	11	16	16	10	9	14	7
	100%	100%	100%	100%	100%	100%	100%	100%	100%

NOTE: Undecided and other candidates are excluded from the table. Own Preference is the aggregate of individual opinions of the panel. Perceived Opinion is the mean of respondents' judgments of what proportion each candidate had at the time of interview (September only). Perceived Outcome is the mean proportion the respondents thought each candidate would receive in November (asked in September and October). Poll data are the aggregated individual opinions of national samples reported in the Gallup polls closest in time to the September and October interviews. Own vote is the aggregate of panel respondents' votes as reported after the election. Actual vote is the national vote in November for the three candidates.

tions of what public opinion would be in November (Perceived Outcome) (Table 4). Respondents' judgments of the outcome differed from the aggregate of their own preferences by an average of 11 percent in September and 5 percent in October. At both time points, the perceived outcome for Anderson was much less than the aggregation of own preferences (11 percent vs. 27 percent, 10 percent vs. 16 percent), while the perceptions of the vote for Reagan were higher than the aggregated preferences of the panel (43 percent vs. 32 percent, 45 percent vs. 36 percent).

These findings are especially intriguing because the sample represents an area known for its liberal viewpoints and highly educated population. The fact that such respondents could reflect the national poll trend so well in their perceptions of the outcome of the election race, and the fact that the expressions of preference could begin to slide in the direction of these perceptions does suggest a direct and perhaps very powerful influence of perceptions of opinions on expressions of own opinion and on actual behavioral outcomes.

References

Allport, Floyd
1937 "Toward a science of public opinion." Public Opinion Quarterly 1:7–23.
Almond, Gabriel
1950 The American People and Foreign Policy. New York: Harcourt, Brace.
Glynn, Carroll, and Jack McLeod
1982 "Public opinion, communication processes and voting decisions." Pp. 759–74 in Michael Burgoon (ed.), Communication Yearbook 6. International Communication Association. Beverly Hills: Sage.
In press "Implications of the spiral of silence theory for communication and public opinion research." In Dan Nimmo and Keith Sanders (eds.), Political Communication Yearbook Vol. 1. Carbondale: Southern Illinois University Press.
Noelle-Neuman, Elisabeth
1974 "The spiral of silence: a theory of public opinion." Journal of Communication 24 (2):43–51.
1977 "Turbulences in the climate of opinion: methodological applications of the spiral of silence theory." Public Opinion Quarterly 41:143–58.
Page, Benjamin, and Robert Shapiro
1982 "Changes in Americans' policy preferences, 1935–1979." Public Opinion Quarterly 46:24–42.
Taylor, D. Garth
1982 "Pluralistic ignorance and the spiral of silence." Public Opinion Quarterly 46: 311–35.

Constituency, Party, and Representation in Congress

BENJAMIN I. PAGE, ROBERT Y. SHAPIRO,
PAUL W. GRONKE, AND ROBERT M. ROSENBERG

EDMUND BURKE, James Madison, J. S. Mill, and others have expressed varying views on how a legislator ought to represent his constituency. Theorists have also offered different ideas about how legislators actually behave. Some, including Downsian theorists working with an "economic theory of democracy," expect representatives to act exactly in accord with the policy preferences of their constituents (first section of Downs, 1957). Others argue that legislators are largely free of popular control and are influenced instead by interest groups' wishes, party loyalties, peer pressures, or their own judgments.

The pioneering Miller and Stokes study of 1958 made possible for the first time a systematic empirical examination of linkages between sampled public opinion and roll call voting in Congress (Miller and Stokes,

Abstract Using congressional districts as primary sampling units, the 1978 National Election Survey provides improved (though still imperfect) measures of district opinion. Together with Census data on district demography, roll call voting scales, and information on congressmen's party and personal characteristics, they permit a new examination of representation in Congress. Using these data we found a high degree of representation of district opinion on social welfare and (surprisingly) on women's issues, nearly as much on racial issues, and much less on law and order or on abortion. District demography and congressmen's party add substantially to the explanation of roll call votes. There is not, however, much "responsible party" representation in Congress. Future representation studies must face questions about the complex interplay among these factors, including reciprocal influences.

Benjamin I. Page holds the Frank C. Erwin, Jr., Chair in Government at the University of Texas, Austin, and is a Research Associate at the National Opinion Research Center. Robert Y. Shapiro is Assistant Professor of Political Science at Columbia University and an NORC Research Associate. Paul W. Gronke is a doctoral student at the University of Michigan. Robert M. Rosenberg is Deputy Director of the State of Florida Division of Pari-Mutual Wagering. The authors wish to thank Calvin Jones for collaborating in this research and Eric Smith, Bruce Peterson, Robert Erikson, Lutz Erbring, Eleanor Singer, and the anonymous reviewers for offering helpful comments. Kathy Bawn provided capable assistance. The research was supported by the National Opinion Research Center and the Social Science Division of the University of Chicago. The 1978 American National Election Study data were made available by the Inter-university Consortium for Political and Social Research; the responsibility for analysis and interpretation is the authors'.

Public Opinion Quarterly Vol. 48:741–756 © by the Trustees of Columbia University
Published by Elsevier Science Publishing Co., Inc. 0033-362X/84/0048–741/$2.50

1963, 1966). A number of researchers using the 1958 data (Cnudde and McCrone, 1966; Achen, 1977, 1978; Erikson, 1978; Karps and Eulau, 1978; Weissberg, 1976, 1978) and other National Election Studies (Stone, 1977, 1979) have conducted increasingly sophisticated empirical analyses, but they have always been limited in their conclusiveness because of poor samples within congressional districts. A small number of interviews was conducted within each district, and the sampling scheme was not designed to produce random samples of opinion for districts.

It is now possible to study representation in Congress using better data on constituents' opinions from the 1978 National Election study (see Erikson, 1981). The 1978 survey, unlike others before, designated congressional districts as primary sampling units (PSUs). These PSUs promise considerable improvement in the representativeness of constituency samples. Moreover, the political times have changed in the more than two decades since 1958. New policy issues have arisen while some old ones have changed or faded. The newer survey, with a better sampling frame, offers a fresh opportunity to examine representation.

At the same time, however, we must not exaggerate the virtues of the 1978 data. For reasons of economy, interview administration within districts was not random; interviews were geographically clustered, counteracting some of the advantages of the new PSUs. Sample sizes (average $n = 21$) in the 108 district PSUs were larger than in 1958 (average $n = 13$), but by no means enormous. In addition, the lack of any parallel survey of congressmen means that some intervening processes involving representatives' attitudes and their perceptions of district opinion cannot be analyzed directly.

Methods

We took the 108 sampled districts as our units of analysis. For each district we computed constituents' mean policy preferences on 10 selected issue items from the 1978 survey, covering social welfare, racial issues, law and order (or civil liberties), and the new issues of women's rights and abortion. In addition, we coded Census data for districts from the *Congressional District Data Book* (U.S. Bureau of the Census, 1973, 1974a, 1974b, 1974c), including median income and education, mean household size, and the proportions of the population that were urban, living in an SMSA, blue collar, black, and of foreign stock. These same district characteristics were also calculated from NES survey respondents' reports, so we could appraise the representativeness of the district samples. Further, we used the Census demographics and individual-level survey data to provide additional measures of constituency attitudes by simulating district-level opinion, using a regression method similar to Erikson's (1978; see also Weber, et al., 1972–73; cf. Seidman, 1975).

Our chief dependent variables were a series of scales of congressmen's roll call votes in the 95th Congress (1977–78), on the issues for which we had measures of constituency policy preferences.[1] Along with the roll call scale scores for each congressman, we also included in our file the National Election Study supplementary data for district votes in 1978 and for congressmen's personal characteristics.

Analysis and Findings

Early in our analysis we found that the 1978 constituency survey data are indeed better than those of 1958. The survey-measured demographic characteristics of districts correlate moderately well—and in some cases quite highly—with the Census figures (cf. Erikson, 1981). Practically all the correlations are higher than those found for the 1958 data (see Erikson, 1978: 518); the average correlation is .69 for 1978, compared with .56 for 1958. This is particularly encouraging in that the 1958 sample was compared with 1960 Census data, whereas our correlations are probably lessened by comparing the 1978 sample with the more distant 1970 Census.[2]

The data are still far from perfect, however. On the average no more than half of the variance in the presumably accurate Census figures can be accounted for by the surveyed district characteristics. This indicates that the 1978 samples for districts are (as the 1978 National Election Study codebook clearly warns) often far off the mark. It is hoped that district samples will be larger and selected more randomly sometime in the future, perhaps through telephone interviewing.

---

[1] To construct roll call measures, we used a method similar to the one described by Clausen and Horn (1977). Eliminating roll calls with a lopsided (90 percent–10 percent or worse) split or with strictly procedural content, we first classified votes into seven issue categories determined by the manifest content of the 1978 survey questions. We then factor analyzed separately the votes within each issue group to identify, if possible, a unidimensional structure. After performing a principal components analysis, we eliminated all votes that loaded on the first factor (unrotated) at less than .6, and excluded some minor redundant roll calls. We were left with 50 economic and social welfare votes (the first factor accounting for 55 percent of the variance), 19 law and order votes (54 percent of the variance), 33 racial votes (57 percent), 14 medical care votes (56 percent), 6 women's rights votes (68 percent), 17 abortion votes (78 percent), and 44 "other liberalism" votes (56 percent).

We constructed two types of roll call scales for each group of votes: factor score indices and simple additive indices (the proportion of liberal votes cast). The average correlation between the two was .98, so that they produce virtually identical results; those reported here are for the additive scales.

[2] Comparing the Census data with the district sample estimates for several population characteristics, we found the following Pearson correlations: percent urban, .77 in 1978 versus .71 in 1958 reported by Erikson (1978:516); percent voting Democratic, .88 versus .60; percent foreign stock, .80 versus .84; percent black, .84 versus .72 (percent nonwhite); percent blue collar, .47 versus .11; mean household size, .51 versus .37; percent living in SMSA, .72 (not available for 1958); median education, .58 (not available for 1958); median income, .61 (not available for 1958).

Nonetheless, the improved district data justify attention to substantive results, some of which are summarized in Table 1. Just as Erikson did in reexamining the Miller-Stokes data, we analyzed the relationships between roll call votes and *simulated* opinions as well as *surveyed* district opinion. Although we are mindful of the perils of the correlation coefficient (Achen, 1977), we were not able to measure "centrism" as Achen (1978) has operationalized it, because no data were available on congressmen's attitudes. Given the different metrics involved, we consider correlations more easily interpretable than unstandardized coefficients as rough measures of "responsiveness."

Using this simple bivariate measure, we found only modest relationships between constituents' policy preferences and the votes of their congressmen. Many of the correlations between roll calls and surveyed opinion fall in the .3−.4 range. The fact that district samples are imperfect suggests that the true relationships are stronger than those reported in the first column of the table, but the correlations between congressional roll calls and the allegedly more accurate simulated district opin-

Table 1. Correlations of District Opinions with Their Congressmen's Roll Call Votes (n = 108)

Issue	*Surveyed Opinion*	*Simulated Opinion*	*Both Surveyed and Simulated*
Social Welfare			.60
Job guarantee	.33	.32	
Medical insurance	.45	.41	
Other liberalism	.43	.43	
Racial issues			.46
Integration—is it govt's business?	.31	.22	
Integration	.26	.35	
Aid to minorities	.41	.24	
Women's rights			.59
Women's rights	.33	.39	
ERA	.54	.41	
Law and order			
Rights of the accused	.04*	.16*	.17*
Abortion			
Abortion	.11*	.05*	.11*
Average for all issues	.32	.30	.39

NOTE: Entries in the first two columns are bivariate r's; in the third column, multiple R's.

* Not statistically significant at the .05 level.

ions are about the same as those for surveyed preferences.[3] (As we will see, the two sets of variables taken together do much better.)

The original Miller and Stokes articles reported a substantial correlation between district opinion and roll calls for racial issues, a moderate correlation for social welfare policies, and a negligible one for foreign policy; they suggested some interesting possible reasons for the differences (Miller and Stokes, 1963, 1966; Miller, 1964). While these findings have been disputed by Achen (1977, 1978), Erikson (1978), and others on methodological grounds, they remain part of the conventional wisdom and a prime topic for investigation. The appearance of new issues on the agenda raises particularly intriguing questions as to where they fit.

In Table 1, our policy questions were grouped a priori (on the basis of their manifest content) into five categories: social welfare, racial issues, women's rights, law and order, and abortion. The similarities and differences across issues are interesting. The Miller and Stokes contrast between social welfare and race issues is erased or reversed. The correlations for job guarantee, medical insurance, and other liberalism are every bit as large—in fact generally larger—than those for aid to minorities and the two integration questions. (It is noteworthy that the highest correlating racial issue, aid to minorities, has a social welfare component.) Furthermore, the new issues of women's rights and the ERA display the same fairly high level of correlations. These genuinely new issues, not present in the 1958 data nor prominent in the political agenda at that time, show representational relationships in the 1970s similar to those of the old issues.

The similarities among issues extend further. Congressmen tended to cast consistently liberal or consistently conservative votes across most of these issues. All the roll call voting indices intercorrelated in the range of .9 (.8 for women's rights), except abortion (.5). Most were also related to the congressmen's political party (average r = .5), a matter which will be discussed further. Thus many issues, including the new ones of women's rights, appear to have been assimilated into a broad liberal-conservative voting pattern in Congress (much clearer than the modest liberal-conservative dimension found in the electorate); and representation, as measured by constituency/roll call correlations, seems much the same across several issues.

---

[3] Part of the difference in this regard between our simulation results and those of Erikson (1978:523) occurred because we did not include the district's party vote in simulating district policy preferences. We feel that the inclusion of the district's party balance (highly correlated, of course, with the congressman's party affiliation) inflates the apparent opinion-roll call relationship when district partisanship or other factors rather than policy preferences may be at work. Our opinion-roll call correlations averaged .45 with party vote included in the simulations and .30 without it.

District-level opinions were simulated from the district Census data using equations of the form:
$\hat{Y} = b_1X_1 + b_2X_2 + b_3X_3 + b_4X_4 + b_5X_5 + b_6X_6 + c$, *where* $\hat{Y}$ = the predicted district-level policy preference on a particular issue, X_1 = median education, X_2 = median income, X_3 = proportion of residents of foreign stock, X_4 = proportion blue collar, X_5 = proportion black, X_6 = proportion urban, and where $b_1, b_2, b_3, b_4, b_5, b_6$, and c are coefficients obtained from the corresponding OLS regression equations estimated from the *individual-level* survey data.

Two issues which do not fit this schema are rights of the accused and abortion. For both of these, the correlations between roll calls and district preferences are very slight. Part of the explanation is that the survey responses for these issues do not correlate very highly at the individual citizen level with most of the other opinion items.[4] Rights of the accused appear to be assimilated by liberalism at the elite but not the mass level. Further, congressional voting on abortion has the weakest relationship with roll call voting on other issues (factor analysis showed it to constitute a strong separate dimension), and it is less strongly related to the congressmen's party ($r = .32$, $p < .05$). Although abortion opinions at the individual level are related to sentiments concerning women's rights and the ERA,[5] it is the only issue which does not fit neatly into either congressmen's or citizens' liberal-conservative orientation. The abortion issue, with its tangled mixture of feminism, personal liberty, lifestyle, and morality, is different from other issues.

Multivariate Analysis of the Representation Process

Thus far we have reported simple bivariate relationships between measured constituency opinion and roll call voting. But this tells us little about the processes by which representation occurs, and it may even conceal the extent of representation. We can learn more by including additional independent and intervening variables in the analysis, pertaining to district demography, congressmen's party affiliations, and congressmen's personal characteristics (e.g., see Froman 1963; Turner and Schneier, 1970).

Our approach to the many possible relationships among these factors has been step by step, examining them one at a time. We begin by simply returning to the variables of constituency opinion and roll call votes. Observing that the surveyed and simulated opinion measures each provide only imperfect but complementary indications of actual district

---

[4] The average correlations across individuals with opinions on all other issue are as follows: for guaranteed job/standard of living, .28; government health insurance, .31; general liberalism, .40; aid to minorities, .40; integration-government's business, .27; integration, .22; women's rights, .30; ERA, .34; rights of the accused, .17; abortion, .17.

[5] Citizen's opinions on abortion and women's issues load in the .4 to .6 range on an oblique factor upon which no other issues except "other liberalism" load as high as .2. When opinions are aggregated to the district level, a stronger women's factor emerges, with abortion loading at .77.

Table 2. Proportion of Variance in Congressmen's Roll Call Voting Explained by District Opinion and Other Factors

	Explanatory Variables			
Issues	*Surveyed and Simulated District Opinion*	*Plus District Demography*	*Plus Congressman's Party*	*Plus Congressman's Personal Characteristics*
Social welfare	.36 (.32)	.40 (.34)	.68 (.65)	.71 (.65)
Racial issues	.21 (.16)	.40 (.35)	.60 (.56)	.64 (.57)
Women's rights	.35 (.33)	.43 (.38)	.51 (.47)	.56 (.48)
Law and order	.03* (.01)	.31 (.26)	.59 (.56)	.63 (.57)
Abortion	.01* (.00)	.18 (.12)	.26 (.20)	.46 (.37)

NOTE: Entries are R^2's (adjusted R^2's in parentheses).
* Not statistically significant at the .05 level.

preferences, we include *both* together in a regression equation predicting roll call votes. The first column of Table 2 presents the proportions of variance explained (R^2) by both surveyed and simulated opinion.[6] For purposes of comparison with the earlier bivariate results, the corresponding multiple R's are displayed in the third column of Table 1. We find some substantial relationships that are notably stronger than the bivariate correlations presented earlier, indicating a very considerable degree of representation.

---

[6] A single social welfare roll call scale was constructed by averaging the measures for economic and social welfare votes, medical care votes, and other liberalism roll call votes, and it is used hereafter rather than the three separate (highly intercorrelated) scales. The roll call scales for social welfare, race, and women's rights votes therefore encompass issues addressed by more than one survey question. For these issues, district opinions on all relevant survey questions are included as separate independent variables in the regressions.

Here and subsequently we report findings mainly in terms of proportions of variance explained rather than coefficients for the effects of individual variables. We wish to summarize the explanatory power of different *sets* of variables. Normally we would have preferred presenting the individual coefficients as well, but this would have been cumbersome and perhaps also confusing, since multi-collinearity makes some of them quite unstable.

It is not usually appropriate to rely on standardized statistics, such as proportion of variance explained or standardized regression coefficients, to estimate the relative effects of variables across equations (or populations; see especially Achen, 1982:69–77), because of sensitivity to differences in the variances of individual variables from one equation to another. For example, if there is less variance in the dependent variable in one equation than in a second equation, a given independent variable can have the same *effect* in each equation (as estimated by an unstandardized coefficient) but explain different proportions of the variance. In our data, however, the standardized and unstandardized coefficients have approximately the same relative magnitudes, because the variances of most of the independent variables (district demographic characteristics, congressman's party, and his other personal characteristics) are identical across equations—they are the same measures for the same districts at the same point in time—and because the several dependent variables, which are all proportions of roll call votes, turn out to have nearly the same variances as well. Because standardized measures produce correct inferences in this case and because there is no convenient, single unstandardized coefficient available to summarize the effects of sets of variables (cf. Heise's (1972) "sheaf coefficient," which is standardized), we find it useful to compare proportions of explained variance. (See also Nie, et al., 1980:303, and footnote 8). When we mention regression coefficients for certain individual variables, however, we will report both the unstandardized (b) and standardized (beta) estimates.

By this measure, the extent of representation appears to be greater than previously thought. Again the differences and similarities among issues are interesting. The most representation, with multiple correlations of approximately .6 (R^2 = .36), appears to occur on social welfare policy and women's rights, a bit less for racial issues, and much less for law and order or abortion. The somewhat greater representation on social welfare than civil rights, while contrary to Miller and Stokes' reports from the 1958 data, is consistent with Erikson's (1978) reexamination of those data looking only at districts outside the South. (By 1978 regional differences on civil rights attitudes had moderated greatly.[7] In simulating opinion from district demographic data, we encountered no more difficulty for the South than for the Non-south.) It may well be, then, that the higher correlation for racial issues originally reported from the 1958 data was a methodological artifact. We would emphasize the similar extent of representation on the issues of social welfare, women, and race.

The very small relationships for law and order and abortion again can be attributed to the fact that they stand somewhat outside the liberal-conservative organization of political attitudes among constituencies which is linked with a strong liberal-conservative voting pattern in the House of Representatives.

Our multiple measures of constituency opinion appear to have eliminated most of the usual attenuation due to sampling error that depresses estimates of the extent of representation. Erikson (1981) adjusted for reliability his bivariate correlations based on 1978 survey-measured opinion (using ADA and ACA liberalism and conservatism scores rather than issue-specific roll-call scales as dependent variables). All the multiple *R*'s reported in Table 1, except for racial issues, are at least as high as his adjusted bivariate correlations.[8] It is, of course, possible that some attenuation remains, because of response errors in the survey data or other factors like the limited predictive power of the simulation equations. Even so, our estimates of the degree of representation are considerably higher than most past estimates.

---

[7] The average correlation was −.09 (not statistically significant at the .05 level) between region (South) and district opinions on the three racial issues questions, and −.14 (n.s.) for the three social welfare items. Region independently explained only 1 percent of the variance (n.s.) in roll call votes on both issues.

[8] The similarity of our findings also tends to vindicate Erikson's use of group ratings to assess representation on the general liberal-conservative dimension, if not on specific issues. The average correlation between our roll call scales (excluding abortion) and the ADA and ACA rating was .87.

The concept of representation underlying these findings is a very particular and perhaps narrow one, involving correspondences between roll call votes and the subjective policy preferences (as revealed by verbal survey responses, sampled and simulated) of constituencies. A broader conception of representation might take into account constituency preferences that are not articulated, and even underlying constituency interests or needs, some of which may be captured, if not determined, by the demographic composition of district populations.[9]

We estimated the extent of this broader kind of representation by adding a number of demographic indicators (median income, median education, proportion black, proportion foreign stock, proportion urban, and proportion blue collar) to the equation predicting roll call votes from district opinion.[10] The results are shown in the second column of Table 2.

We found that district demographic characteristics do indeed explain additional variation in congressional voting, raising significantly ($p < .05$) the proportion of variance accounted for to nearly one half on some issues. The great increase in explanation of law and order votes is attributable mainly to the fact that congressmen from highly urban districts (unstandardized partial $b = .39$, beta $= .27$, $p < .05$) and districts with a high proportion of foreign born or first generation residents ($b = 1.20$, beta $= .44$, $p < .05$) voted more liberally than was indicated by district opinions on the "rights of the accused" question. Similarly, congressmen from these types of districts voted more pro-minority (b's of .36 and 1.08, $p < .05$, betas of .24 and .38, $p < .05$, respectively) on racial matters. Congressmen from blue-collar districts voted more frequently against abortion ($b = -3.31$, beta $= -.23$, $p < .05$) and those from black districts more for abortion ($b = .66$, beta $= .24$, $p < .05$) than their constituents' articulated preferences would predict.

Taken together these results suggest that there may be a substantial amount of contextual or demographic representation, over and above the representation of articulated policy preferences. Assuming that it does not merely reflect better opinion measurement, this effect of demographics could signify representation of unarticulated needs, or merely demographically linked interest group influence or party loyalty overriding constituency views.

---

[9] Of course the concept of representation can be expanded much further. See Eulau and Karps (1978) and the essays in Jewell and Leowenberg (1979).

[10] We had to exclude the simulated opinion measures from these regressions, because they are linear combinations of, and therefore perfectly collinear with, the separate demographic indicators. Note that the separate demographic variables may be better predictors of preferences than their fixed linear combinations, so that a part—we would judge a small part—of the improvement in predicting roll calls may be attributable to representation of policy preferences rather than demography, especially on the single-item issues of abortion and law and order.

By this measure, differences among issues are reduced: there is approximately the same total amount of representation for social welfare, racial, and women's rights issues. There is less, though no longer so much less, on law and order and abortion, where representation occurs very heavily through demographic factors rather than explicit constituency opinions.

We next examined the role of the incumbent congressman's party affiliation. The third column of Table 2 shows the amount of variance in roll call votes attributable to constituency opinion, demographic composition, and the representative's party, all taken together. Clearly, party adds a great deal. For the first time we are accounting for more than half the variance in roll call voting on most issues.

In order to learn more about *processes* of representation, we decomposed the explained variance into portions attributable to constituency alone, to congressman's party alone, or shared by both. These are reported in Table 3. As shown in the third column, a congressman's party is a powerful predictor in itself. Independently of constituency opinion or demography, political party explains 28 percent of the variance in social welfare votes, 28 percent also for law and order, and 20 percent for voting on racial issues. These independent effects of party could reflect support for party positions, including those of party leadership of the president, or they could simply result from the distinctive policy attitudes that McClosky (1960) and others have found among party elites. Given the decentralized structure of Congress, shared ideology seems more plausible than party discipline. Party voting in Congress is clearly linked to

Table 3. Partitioning the Variance in Roll Calls Explained by Congressman's Party, Constituency Opinion, and Demographic Characteristics

	Proportion of Variance Explained By:			
Issues	*Party, Opinion, and Demographic Characteristics Together*	*Opinion and Demographic Characteristics Alone*	*Party Alone*	*Opinion/Demographic and Party, Shared*
Social welfare	.68 (.65)	.28 (.24)	.28 (.30)	.12 (.11)
Racial issues	.60 (.56)	.32 (.28)	.20 (.21)	.08 (.07)
Women's rights	.51 (.47)	.41 (.37)	.08 (.09)	.02* (.01)
Law and order	.59 (.56)	.25 (.21)	.28 (.29)	.06 (.06)
Abortion	.26 (.20)	.16 (.10)	.08 (.08)	.02* (.02)

NOTE: Figures in parentheses are adjusted R^2's.
* *Not* statistically significant at the .05 level.

constituents' party loyalties, which sometimes have historical roots unrelated to policy preferences or demographic characteristics for which we have data.[11]

It is also apparent from the second column of Table 3 that constituency opinion and demography have a strong impact on roll calls quite independent of the congressman's party. Some of this impact results from the kind of demographic representation we have alluded to, and some of it seems to work through classical processes of representation in which districts elect congressmen who share their policy preferences or who are willing to act as delegates in representing them.

We consider the variance in roll calls that is shared between constituency factors and congressman's party to be an indicator of the extent to which the political parties are acting "responsibly" to effect representation in Congress. It reveals the degree to which partisan voting by congressmen reflects differing constituencies of the two parties. Particularly given the liberal-conservative division in Congress and (to a lesser extent) among the citizenry, and the historical cleavages between Democrats and Republicans along liberal-conservative lines, we expected to find a substantial amount of this responsible party kind of representation.

Somewhat to our surprise, however, party and constituency characteristics shared relatively little of the explained variance in roll calls (see the fourth column of Table 3). Most of the effects of party and constituency are independent of each other. There is some responsible party representation (12 percent of the variance) on social welfare votes, the policy area most closely tied to the New Deal party alignment; but there is hardly any on the more recent and less party-related concerns of women's rights and abortion. Overall, representation occurs mostly independently of party, and party has most of its effect separate from constituency opinions or demographic characteristics. To a substantial extent, congressional partisanship is an elite (and perhaps interest group related) phenomenon.

We are convinced that the variables discussed so far capture at least the basic framework of the representation process, and the major determinants of roll call votes. Except for the case of abortion, these variables collectively account for half or more of the variance in roll call voting, and in the case of social welfare, approximately two-thirds of it.[12]

---

[11] On average, district partisanship (measured, imperfectly, by the 1978 congressional party vote) independently accounted for less than 1 percent of the variance (n.s.) in roll call votes, but by itself it predicted the congressmen's party affiliation better than all the other constituency factors combined (independently explaining 38 percent of the variance). We did not include it in our analysis, because it adds virtually nothing to the explanation of congressional voting and because it has a complex and probably reciprocal relationship with constituency policy preferences. The role in representation of district party loyalty deserves further exploration, however.

The attitudes of the representatives themselves and their perceptions of sentiments in their districts are important intervening and, perhaps, independent variables not included in the 1978 data. The congressman's party affiliation may subsume these unmeasured variables to some extent. Personal characteristics of the legislator may do so as well; they may intervene in demographic as well as attitudinal processes of representation, and may have an independent impact of their own on roll call voting. Accordingly, we added a final set of variables, congressmen's personal characteristics (race, religion, occupational background, seniority, age, and sex) to the regressions predicting roll call votes. The results are displayed in the last column of Table 2.

On most issues congressmen's personal characteristics do not directly add very much to what has gone before; they explain only 3 percent to 5 percent (not statistically significant at the .05 level) additional variance in roll calls. They are more important as intervening variables—perhaps standing for congressmen's policy preferences—that are influenced by constituency preferences and demographic characteristics, and are related to congressman's party.

On the abortion issue, however, congressmen's characteristics make a substantial difference ($p < .05$), nearly doubling the variance explained. In particular, quite independently of their party or their constituencies' characteristics, Catholic congressmen tend to vote against abortion ($b = -.45$, beta $= -.46$, $p < .05$) and black congressmen to vote for it ($b = .39$, beta $= .23$, $p < .05$). Of all our issues, abortion provides the clearest example of weak constituency representation and of legislators likely to be voting in accordance with their own consciences, interest group pressures, or their ethnic or religious loyalties.[13]

---

[12] Our confidence is bolstered by a bit of collateral evidence. We divided districts according to their responses (above or below the median) to a survey question asking, in effect, whether respondents wanted their congressmen to act as delegates or trustees. (See Eulau, et al., 1959; McCrone and Kuklinski, 1979). We then compared the results of our equations predicting congressmen's votes for the two types of districts. Substantially more of the variation in roll call voting was accounted for among the districts showing the greater desire that congressmen act as delegates: an average R^2 of .66, versus .48. Oddly, however, this difference is attributable almost entirely to the increased impact of congressmen's party and districts' demographic characteristics. There is little or no more representation of measured constituency opinions in the districts showing greater preference for delegates.

[13] Although there are no Census data for the percentage of Catholics in congressional districts, we did examine the proportion of self-reported Catholics in our NES district samples. Using this measure, we found that Catholic congressmen were more likely to be elected from districts with large proportions of Catholics ($r = .39$, $p < .05$). This seems to suggest on the face of it that Catholic congressmen voted in accord with the opinions of their Catholic subconstituencies, which were presumably against abortion; but in fact Catholic

Conclusion

Using multiple measures of constituency opinion—both surveyed and simulated—we found a substantial amount of correspondence between congressmen's roll call votes and their constituencies' policy preferences, with multiple R's reaching .60. This was particularly true of social welfare and the new women's issues and (to a slightly lesser extent) racial issues, all of which fall on a liberal-conservative dimension of citizens' attitudes. The extent of representation is higher than previously reported, though still not accounting for much more than one-third of the variance in roll call votes.

Constituency opinions were much less well represented on the issue of law and order/rights of the accused, which conforms to the liberal-conservative continuum only at the elite level, and on the issue of abortion, which is dissociated from general liberalism-conservatism among congressmen as well as citizens. Representation may work best on issues that are institutionalized in party cleavages and linked to broad ideology among the public, where opinions are fairly firmly held and information about congressmen is easily obtainable.

The demographic characteristics of constituencies add considerably to the explanation of roll call votes, especially on racial, law and order, and abortion issues. Further work is needed to distinguish the extent to which this signifies direct demographic representation of unarticulated needs, as opposed to augmented measurement of policy preferences, or interest group activity, or demographically linked aspects of party loyalty.

The congressman's party affiliation is a strong predictor of roll call votes on most issues. But only to a surprisingly small extent does it serve, "responsible party" style, as an intervening variable for constituency opinion or demography. To a very great extent, congressmen's party acts as an independent factor on the elite level, probably reflecting shared ideology and/or interest group coalitions.

Finally, congressmen's personal characteristics do not have much independent effect on roll call votes, except in the case of abortion, where Catholic congressmen vote more anti-abortion (and black congressmen more pro-abortion) than their constituencies' opinions or demography

---

citizens opposed abortion only weakly or not at all ($r = -.13, p < .05$ at the individual level for the national sample, and a contrary .15, n.s., at the district level). Perhaps the Catholic congressmen misperceived constituents' preferences or responded to Catholic-based pressure groups that were important in their districts. But in any case we found that the congressmen's personal characteristics explained 20 percent of the variance in abortion votes, independently of the (estimated) percentage of Catholics in their districts—supporting interpretations based on the congressman's own beliefs or loyalties, or the influences of beyond-district interest groups.

would predict, perhaps for reasons of interest group pressure or personal belief. On most other issues, congressmen's personal characteristics are simply intervening variables, linking their constituents' attitudes and characteristics and party loyalties to legislative votes.

In reporting these findings we do not wish to minimize the remaining uncertainties about representation, or the difficulties that stand in the way of clarifying them. Future studies of representation would certainly profit from larger and better district samples. In order to disentangle different processes of representation, there should also be direct measurement of congressmen's attitudes and perception. But even with such improved data scholars will face some knotty analytical problems.

In this paper we, like Miller and Stokes and most others, have examined roll call votes and constituency policy preferences within a recursive model, postulating one-way causation. There is abundant reason, however, to think that many relevant variables (particularly congressmen's attitudes and perceptions) may affect each other reciprocally. That is, a nonrecursive model is the appropriate one (see Forbes and Tufte, 1968; Asher, 1976; Kuklinski and McCrone, 1980). The estimation of such a model is likely to prove quite difficult because of the lack of theoretically exogenous variables needed for identification. Thus, there is still a great deal to learn about the processes by which representation in Congress occurs, and no lack of difficulty in learning it.

References

Achen, Christopher

1977 "Measuring representation: perils of the correlation coefficient." American Journal of Political Science 21:805–15.

1978 "Measuring representation." American Journal of Political Science 22:475–510.

1982 Interpreting and Using Regression. Beverly Hills, CA: Sage Publications.

Asher, Herbert B.

1976 Causal Modeling. Beverly Hills, CA: Sage Publications.

Clausen, Aage, and Carl Van Horn

1977 "How to analyze too many roll calls and related issues in dimensional analysis." Political Methodology 4:313–31.

Cnudde, Charles F., and Donald J. McCrone

1966 "The linkage between constituence attitudes and congressional voting behavior: a causal model." American Political Science Review 60:66–72.

Downs, Anthony

1957 An Economic Theory of Democracy. New York: Harper.

Erikson, Robert S.

1978 "Constituency opinion and congressional behavior: a reexamination of the Miller-Stokes data." American Journal of Political Science [illegible]:511–35.

1981 "Measuring constituency opinion: the 1978 congressional election study." Legislative Studies Quarterly 6:235–45.

Eulau, Heinz, John C. Wahlke, William Buchanan, and Leroy C. Ferguson

1959 "The role of the representative: some empirical observations on the theory of Edmund Burke." American Political Science Review 53:742–56.

Eulau, Heinz, and Paul D. Karps
1977 "The puzzle of representation: specifying components of responsiveness." Legislative Studies Quarterly 2:233–54.

Forbes, Hugh Donald, and Edward R. Tufte
1968 "A note of caution in causal modeling." American Political Science Review 62:1258–64.

Froman, Lewis, Jr.
1963 Congressmen and Their Constituencies. Chicago: Rand McNally.

Heise, David R.
1972 "Employing nominal variables, induced variables, and block variables in path analyses." Sociological Methods and Research 1:147–73.

Jewell, Malcolm E., and Gerhard Loewenberg (eds.)
1979 Legislative Studies Quarterly 4 (November).

Karps, Paul D., and Heinz Eulau
1979 "Policy representation as an emergent: toward a situational analysis." Pp. 207–31 in Heinz Eulau and John Wahlke (eds.), The Politics of Representation. Beverly Hills, CA: Sage Publications.

Kuklinski, James H., and Donald J. McCrone
1980 "Policy salience and the causal structure of representation." American Politics Quarterly 8:139–64.

McClosky, Herbert, Paul J. Hoffman, and Rosemary O'Hara
1960 "Issue conflicts and consensus among party leaders and followers." American Political Science Review 54:406–27.

McCrone, Donald J., and James H. Kuklinski
1979 "The delegate theory of representation." American Journal of Political Science 23:278–300.

Miller, Warren E.
1964 "Majority rule and the representative system of government." Pp. 343–76 in Erik Allardt and Yrgo Littunen (eds.), Cleavages, Ideologies and Party Systems. Helsinki: Academic Bookstore.

Miller, Warren E., and Donald E. Stokes
1963 "Constituency influence in Congress." American Political Science Review 57: 45–56.
1966 "Constituency influence in Congress." Pp. 351–73 in Angus Campell, Philip E. Converse, Warren E. Miller, and Donald E. Stokes (eds.), Elections and the Political Order. New York: John Wiley and Sons.

Nie, Norman H., Sidney Verba, and John R. Petrocik
1980 The Changing American Voter, enlarged ed. Cambridge, MA: Harvard University Press.

Seidman, David
1975 "Simulation of public opinion: a caveat." Public Opinion Quarterly 39:331–42.

Stone, Walter J.
1977 "A panel analysis of representation in congress: a preliminary report." Paper, Annual Meeting of the American Political Science Association, Washington, D.C.
1977 "Measuring constituency-representative linkages: problems and prospects." Legislative Studies Quarterly 4:623–39.

Turner, Julius, and Edward V. Schneier, Jr.
1970 Party and Constituency: Pressures on Congress. Baltimore: Johns Hopkins Press.

U.S. Bureau of the Census
1973 Congressional District Data Book, 93rd Congress. Washington, D.C.: United States Government Printing Office.
1974a Congressional District Data Book, 94th Congress, Texas. Washington, D.C.: U.S. GPO.
1974b Congressional District Data Book, 94th Congress, California. Washington, D.C.: U.S. GPO

1974c Congressional District Data Book, 94th Congress, New York. Washington, D.C.: U.S. GPO.

Weber, Ronald E., Ann H. Hopkins, Michael L. Mezey, and Frank J. Munger

1973 "Computer simulation of state electorates." Public Opinion Quarterly 36:549-65.

Weissberg, Robert

1976 Public Opinion and Popular Government. Englewood Cliffs, N.J.: Prentice Hall, Inc.

1978 "Collective vs. dydatic representation in congress." American Political Science Review 72:535-47.

Alternative Question Forms for Attitude Scale Questions in Telephone Interviews

PETER V. MILLER

THE MARKED growth in the use of the telephone for survey interviewing has spurred a number of queries about the effects of that medium on data quality. These concerns generally have involved the comparability of results obtained in telephone and face-to-face interviews; a related issue is whether survey techniques developed for face-to-face interviews can be utilized in telephone contacts. For example, can items which employ numerical scales be presented to telephone respondents without the visual aids commonly accompanying them in person interviews?

As a contribution to this debate, this paper presents the findings of a national study which experimentally compared two methods of presenting seven-point scale attitude questions to telephone respondents. One approach simply presented the response categories in the questions

Abstract The absence of visual aids in telephone interviews has led investigators to seek methods for asking questions which are commonly accompanied by aids in personal interviews. This study experimentally compared two approaches to asking seven-point scale attitude questions in a national ($N = 4{,}300$) telephone health survey. A random half-sample received the attitude questions as single-step numerical selection tasks. The questions were administered to the other half-sample in two stages, the first asking for a general verbal statement of attitude and the second asking for a more detailed specification. The two question forms produced comparable results in univariate distributions. Items administered with the one-step form showed slightly higher intercorrelations. This study represents one step in the process of empirically testing data collection procedures which have been deemed necessary in the rich folklore of survey research. Overall, the results show more comparability than might be anticipated, given the multiple differences between the forms.

Peter V. Miller is an Associate Professor in the Department of Communication Studies at Northwestern University. This is a revised version of a paper presented at the annual meeting of the American Statistical Association, August 1981. Data for the study were collected when the author was a Study Director at the Survey Research Center, University of Michigan. The author wishes to thank Robert Groves, Duane Alwin, Howard Schuman, and Charles Cannell for help and advice. The author alone is responsible for any errors. The project was funded by the National Center for Health Statistics (contract 233-78-2034). The contents of this paper do not necessarily represent the views of the National Center for Health Statistics.

Public Opinion Quarterly Vol. 48:766–778 © by the Trustees of Columbia University
Published by Elsevier Science Publishing Co., Inc. 0033-362X/84/0048–766/$2.50

through verbal description. The alternative method presented the categories in two stages, asking respondents first for a general statement of affect and following that with a more specific query in order to place the respondent on the scale. The two-step format is a common question adaptation used to compensate for the presumed difficulty of responding to an item with many scale points without the aid of a visual depiction of the scale. Judgments concerning the method's utility have not been empirically based, however. This study sought to bring data to bear on the issue of whether the adaptation yields different results from the simpler, one-step approach and whether a preference for either technique could be justified empirically. To address these questions, we compare univariate and bivariate statistics and missing data for questions employing the different formats.

Visual Aids in Personal Interviews

Visual aids, often called "show cards," are used in personal interviews for a variety of reasons: as aids to recall of events or behavior (e.g., in readership studies), as devices to reduce question threat (e.g., in income measurement), and as "prompts" to remind the respondent of response categories. This study is concerned with the third type of usage. How should one present questions which seem to require visual "prompts" to respondents?

The necessity of using such visual aids in personal interviews has not been tested experimentally, but they are a common feature of survey practice and are suggested in research textbooks (see, e.g., Moser and Kalton, 1972; Gorden, 1975; Bailey, 1978). Questions which offer a number of *nominal* response categories may require the assistance of show cards more than those which involve *ordered* response alternatives. For example, Kohn's (1977) measurement of parental values involves selection and ranking tasks using a list of 13 child characteristics. A visual aid seems to be mandated by such questions, since without it the interviewer may have considerable difficulty communicating response alternatives. Questions involving ordered response categories may not present as great a problem, since the respondent may be aided in answering by a knowledge of the continuum underlying the choices. Still, as the number of ordered categories gets larger, the difficulty in remembering distinctions among individual scale points increases. While there is no established "magic number" at which point the task of coping with response categories without a visual aid is believed to become difficult, it is common for investigators to use show cards in personal interviews with scales involving as few as five points, particularly if several such questions are presented in succession. Interviewers often strongly prefer to use show cards in such situations rather than to read response alternatives several times, and this preference encourages investigators to use a visual aid whether respondents actually need one or not. Rogers (1976) noted, for example, that telephone interviewers in her study comparing telephone and personal modes of data collection disliked asking a set of 15 five-point scale questions because they had to repeat response alternatives frequently, but that respondents appeared to cope well.

Adapting "Visual Aid" Questions for the Telephone

When faced with the necessity of administering in a telephone interview a rating scale question which seems to require a show card for adequate comprehension, investigators have adopted several different strategies. Apart from the simple approach of having interviewers muddle through without the visual aid, repeating response options frequently, researchers have tried to mail visual aids to respondents prior to the interview (e.g. Durako and McKenna, 1980) or to request that respondents construct their own visual aids for use during the interview (e.g., by listing response alternatives, drawing a scale, or using the telephone dial as a scale representation).

A less burdensome approach for the respondent is to break up the problematic question into two or more simpler queries, each involving fewer response alternatives. This strategy was employed by Groves and Kahn (1979) in their national field experiment comparing telephone and personal interviews. In the telephone portion of their study, these investigators compared seven-point "quality of life" rating scales presented in one or two steps. The findings of their investigation serve as a point of departure for the study described in this paper.

Groves and Kahn found more extremely positive responses with the one-step version of the question than with the two-step version. On the other hand, the two-step version tended to yield more "neutral" and "slightly positive" responses. The inter-item correlations among the quality-of-life items averaged about five points higher for the two-step format. As a whole, the findings suggested that the two-step format, with less extremity bias and higher inter-item correlations, yielded slightly better results than the one-step version.

But a number of complicating factors call this inference into question. First, the two versions of the quality-of-life questions differed not only in format, but also in the response categories which respondents could pick. For each rating of life quality, the one-step scale involved selecting a number between 1 and 7, where the polar categories were "completely satisfied" and "completely dissatisfied." The middle (4) category was labeled "neutral." This is an adaptation of a scale developed by Campbell et al. (1976). The two-step scale, on the other hand, involved selecting among seven verbally labeled categories. In the first stage, the respondent was asked whether he or she was "mostly satisfied," "mostly dissatisfied," or "mixed" in evaluating some aspect of life. The second step involved selecting among "delighted," "pleased," and "mostly satisfied" for those who expressed satisfaction in the first step, or among "terrible," "unhappy," or "mostly dissatisfied" for those who said they were dissatisfied in the first step. This scale is adapted from Andrews and Withey (1976). The two-step question simplified the task of selecting response options by narrowing the field to three choices in each question. Those who replied "mixed" to the first step were not asked to specify further their feelings; for them, the two-step format was effectively one-step. Groves and Kahn note this as a possible reason for a larger number of respondents opting for this category across questions in the two-step version.

These substantive differences between the one- and two-step versions may be partially responsible for the results reviewed earlier. The comparison of scale formats is confounded by differences in question wording. Having selected the "delighted"-"terrible" scale for the two-step version, Groves and Kahn could have compared it to a theoretically "clean" but practically awkward one-step version of the same scale. This option would have entailed having interviewers read the seven scale categories to respondents and, undoubtedly, repeat them several times. (The "delighted"-"terrible" scale is not conducive to conversion into a numerical scale.) The other strategy, adopted by Groves and Kahn, was to compare the "delighted"-"terrible" two-step version to the "satisfaction" scale which could easily be administered in one step as a numerical scale. Such tradeoffs between theoretical "purity" and practical necessity are common in methodological research.

Nevertheless, the Groves and Kahn study left open the question of whether question format, in and of itself, could make a difference in responses to seven-point scale questions administered on the telephone. In addition, because of the way the two-step version was constructed in that study (with no second step for those picking the "mixed" (neutral) category in the first step), the research raised questions about how this format should be designed. As part of a larger study at the Survey Research Center on the feasibility of telephone interviewing for the National Health Interview Survey, research on the effects of question format was designed to address these issues.

Research Design

The SRC-HIS telephone experiment involved a national sample of households selected by RDD methods. Health data from the telephone interviews was compared with information gathered simultaneously from a national sample of respondents interviewed in person by the Bureau of the Census (see Cannell, et al., 1981). In the telephone portion of the study, some 4,300 families were selected, in each of which a family respondent was designated to report on the health of him or herself and that of other adult members of the family.

At the end of the interview, the family respondent was asked a set of five questions about his or her own health and life satisfaction. The respondents were randomly assigned to receive either the one- or two-step version of these questions. The questions asked respondents to rate in succession their satisfaction with their *health and physical condition* in general, their *physical ability* to perform desired activities, their *energy or "pep,"* their *resistance to illness*, and their *life as a whole*. Examples of the two question formats for these items are displayed in Figure 1.

To facilitate the scale format comparison, the "satisfaction" scale used by Groves and Kahn for their one-step version served as the basis for both formats in this study. As can be seen in Figure 1, the one-step version asked for a number between one and seven, with verbal labels attached to the poles of the scale and the midpoint. By contrast, the two-step version asked for verbal responses, and split the response decision into two stages. The first stage of the question asked the respondent to give a general feeling ("satisfied," "dissatisfied," or "in the middle") and then the second step of the question asked for a further specification of that initial statement ("completely satisfied," "mostly satisfied," or "somewhat satisfied"). People who responded "in the middle" to the first query were also probed to see if they were leaning toward satisfaction or dissatisfaction or were truly ambivalent. Thus the operationalization of one- and two-step scales in this study differed from that of Groves and Kahn by basing both formats on a "satisfaction" scale and by having a follow-up question for those selecting the middle category in the two-step version.

These alterations defined the nature of the experimental comparison more narrowly than that earlier study—eliminating the substantive differences in response categories and focusing more on presentation format. Still, there remain a number of features of the two formats which could account for response differences. Most notably, the one-step format asks for a numerical response while the two-step version asks for a verbal response. One could imagine asking respondents to pick a verbal category in both formats, further narrowing the experimental compari-

ONE-STEP: Now, I'll ask you to give me a number between one and seven that describes how you feel about your health. . . "One" stands for completely *dis*satisfied, and "seven" stands for completely *satisfied.* If you are *right in the middle*, answer "four." So, the low numbers indicate that you are *dis*satisfied; the high numbers that you are satisfied.

First, what number comes closest to how satisfied or dissatisfied you are with your *health and physical condition in general*?

_____ Number

8. Never thought; No feelings

TWO-STEP: Now, thinking about your *health and physical condition in general*, would you say you are *satisfied*, *dissatisfied*, or *somewhere in the middle*?

7. Satisfied	1. Dissatisfied
How satisfied are you with your health and physical condition—completely satisfied, mostly, or somewhat?	How dissatisfied are you with your health and physical condition—completely dissatisfied, mostly, or somewhat?
7. Completely	1. Completely
6. Mostly	2. Mostly
5. Somewhat	3. Somewhat

4. In the middle

If you had to choose, would you say you are closer to being *satisfied* or *dissatisfied* with your health and physical condition, or are you *right in the middle*?

5. Satisfied

3. Dissatisfied

4. In the middle

Figure 1. Examples of One- and Two-Step Question Formats

son. This option would entail reading the seven categories to respondents in the one-step version—for example, "Would you say you are completely satisfied, mostly satisfied, somewhat satisfied, right in the middle, somewhat *dis*satisfied, mostly *dis*satisfied or completely *dis*satisfied?" While operationalizing the one-step version in this way would make the experimental comparison conceptually neater, it would also pose some rather large difficulties for interviewers who had to read the numerous choices and who might be forced to repeat the response options frequently to convey them to respondents. Thus, presenting the seven verbal categories to respondents in a single step is an unlikely *field* solution, practically speaking. (Campbell, et al. (1976), in fact, employed the numerical version of the question in their original use of satisfaction scale even though their interviews were conducted *in person.*) The experimental comparison presented here, then, is shaped by the dual goals of comparing question formats while at the same time defining those formats in ways common to survey practice. Response differences may result from the multiple factors which go into making up the two versions.

Analyses

Examination of the findings for this experiment begin with frequency distributions of the satisfaction items by question form. Despite the multiple differences between the question forms noted above, the distributions for the measures seem little affected by the scale format (Table 1). There was a slight tendency for the two-step version to produce higher mean scores across the five questions, although the means are significantly different only in the case of the fourth item, "resistance to illness," despite the large *N*'s in the comparison groups. Higher proportions of respondents receiving the two-step form reported that they were "completely satisfied" with their health and life. It is not clear why this is the case. It may be that since the "completely satisfied' category was presented first in the list in the second stage of the two-step version, respondents were more likely to choose it (Fig. 1). On the other hand, this logic would suggest that respondents receiving the two-step form would have been more likely to select the "completely dissatisfied" category as well, and the data in Table 1 do not show that outcome.

For several of the satisfaction questions, those receiving the one-step version more frequently reported that they were ambivalent or "right in the middle." This finding is exactly the opposite of Groves and Kahn's results, and appears to have come about because of the added follow-up question in the two-step format for those who said they were "in the middle" in the first stage (see Fig. 1). There was also a tendency for those

Table 1. Univariate Frequency Distributions for Five Health Satisfaction Questions by Question Form

	Scale	*Completely Dissatisfied 1*	*%*	*2*	*%*	*3*	*%*	*4*	*%*	*5*	*%*	*6*	*%*	*Completely Satisfied 7*	*%*	*DK/NA*	*Total*	*Mean*	*S.D.*
General health & physical cond.	1-Step	37	(1.7)	36	(1.7)	100	(4.6)	354	(16.3)	475	(21.9)	667	(30.7)	504	(23.2)	49	2221	5.39	1.35
	2-Step	25	(1.2)	45	(2.2)	160	(8.0)	214	(10.7)	400	(20.0)	685	(34.2)	476	(23.7)	80	2084	5.43	1.39
Physical ability	1-Step	50	(2.3)	35	(1.6)	83	(3.8)	299	(15.7)	339	(15.7)	615	(28.5)	737	(34.2)	64	2221	5.61	1.44
	2-Step	38	(1.9)	36	(1.8)	124	(6.2)	193	(9.7)	244	(12.2)	626	(31.3)	736	(36.9)	90	2086	5.70	1.45
Energy or pep	1-Step	64	(3.0)	52	(2.4)	150	(6.9)	422	(19.5)	482	(22.2)	570	(26.3)	427	(19.7)	54	2220	5.13	1.48
	2-Step	48	(2.4)	78	(3.9)	190	(9.5)	359	(18.0)	262	(13.2)	492	(24.7)	563	(28.3)	91	2082	5.23	1.62
Resistance to illness	1-Step	42	(2.0)	28	(1.3)	58	(2.7)	241	(11.2)	312	(14.6)	674	(31.5)	788	(36.8)	79	2221	5.77	1.36
	2-Step	24	(1.2)	28	(1.4)	64	(3.2)	181	(9.2)	188	(9.5)	583	(29.5)	907	(45.9)	91	2076	5.97	1.32
Whole life	1-Step	24	(1.1)	22	(1.0)	54	(2.5)	257	(12.0)	443	(20.7)	721	(33.8)	615	(28.8)	86	2221	5.67	1.24
	2-Step	9	(0.5)	27	(1.4)	73	(3.7)	285	(14.3)	284	(14.3)	633	(31.8)	681	(34.2)	92	2083	5.73	1.28

queried with the one-step version to choose the number 5 category proportionately more often than did those who received the two-step form. This tendency was balanced by the relatively greater frequency with which those responding to the two-step version chose the number 3 category. Finally, the two-step version consistently yielded higher missing data values across the items, as respondents had more opportunity to volunteer "don't know" responses or to opt out of the sequence. None of these differences in individual categories between the two forms are large, by visual inspection, although the distributions for each question format for each item are significantly different by chi-square test. Thus, if one were going to report single categories of response, one would have to take the individual category differences seen here into account.[1]

Following the example of past question form research (e.g., Schuman and Presser, 1981), analyses were undertaken to see if question form differences could be specified by the salience of the topic for respondents. Mean scores for the five questions by question form were examined for respondents who reported thinking about their health "very often" or "often," "now and then," or "rarely" or "never." The findings did not support the hypothesis of a specification of the question form differences, as the mean scores for both versions of the questions were affected similarly by the subjective salience of the health topic (Table 2). Like-

Table 2. Mean Health Satisfaction Scores by Question Form and Perceived Salience of Health

	1-Step Scale			*2-Step Scale*		
	Mean	*S.D.*	*N*	*Mean*	*S.D.*	*N*
Think about health "very often" or "often"						
General health	5.16	1.49	1018	5.14	1.51	910
Physical ability	5.46	1.58	1015	5.53	1.59	902
Energy	5.03	1.57	1017	5.04	1.73	908
Resistance to illness	5.59	1.48	1011	5.82	1.44	894
Whole life	5.59	1.28	1006	5.56	1.37	903
Think about health "now and then"						
General health	5.48	1.16	747	5.54	1.23	718
Physical ability	5.64	1.27	745	5.79	1.27	711
Energy	5.07	1.35	745	5.26	1.49	711
Resistance to illness	5.83	1.20	735	6.02	1.19	711
Whole life	5.63	1.15	737	5.79	1.16	709
Think about health "rarely" or "never"						
General health	5.89	1.21	335	5.97	1.16	327
Physical ability	5.99	1.27	335	5.99	1.31	330
Energy	5.46	1.42	336	5.67	1.49	326
Resistance to illness	6.10	1.24	335	6.28	1.14	323
Whole life	5.94	1.22	328	5.98	1.24	327

---

[1] I am indebted to an anonymous reviewer for noting this *caveat*.

wise, the effects of question form were not specified by whether or not the respondent reported suffering from chronic conditions (Table 3).

The effects of question form on relationships among the satisfaction items is examined in Table 4. We find that the intercorrelations among the items average five to seven points lower for the two-step version of the questions. Comparing the variance-covariance matrices for the set of items in each question form through a confirmatory factor analysis (COFAMM, International Educational Resources, 1976) suggests that the lower correlations found for the two-step scale are due mostly to the larger *variances* in those items (see Table 1). Correlations are equal to covariances divided by the product of the variances, and the COFAMM analyses suggested that the covariances in the two matrices were not significantly different (χ^2 with 10 d.f. = 14.564, $p < 0.15$), while the overall variance-covariance matrices did differ significantly (χ^2 with 15 d.f. = 99.236, $p < 0.001$). The larger variances for items employing the two-step format appear to result from the greater tendency by some respondents queried with that form to select the "completely satisfied" (7) category and others to choose the "somewhat dissatisfied" (3) option. (The latter finding seems to result from the fact that the two-step form probed the "middle" category to see if respondents were leaning toward satisfaction or dissatisfaction.) The answers of those who received the one-step form, on the other hand, tended more to cluster in the three positive categories. The only exception to these patterns is the case of the "resistance to illness" item, which shows a lower variance for the two-step form. Finally, as was the case when examining mean scores, neither health salience nor the experience of chronic conditions specifies the difference between the correlation matrices for the two question forms.

Table 3. Mean Health Satisfaction Scores by Question Form and Chronic Conditions

	1-Step Scale			*2-Step Scale*		
	Mean	*S.D.*	*N*	*Mean*	*S.D.*	*N*
Respondents reporting one or more chronic conditions						
General health	4.77	1.56	705	4.85	1.56	704
Physical ability	4.86	1.72	699	4.94	1.72	709
Energy	4.52	1.63	701	4.73	1.76	705
Resistance to illness	5.28	1.61	691	5.52	1.54	700
Whole life	5.51	1.40	684	5.59	1.34	707
Respondents reporting no chronic conditions						
General health	5.69	1.14	1,468	5.75	1.18	1,301
Physical ability	5.97	1.13	1,459	6.12	1.08	1,288
Energy	5.42	1.30	1,466	5.49	1.47	1,287
Resistance to illness	5.99	1.16	1,452	6.21	1.11	1,275
Whole life	5.74	1.15	1,452	5.79	1.24	1,285

Table 4. Pearson Product-Moment Correlations for Five Health Satisfactions Questions by Question Form

	1-Step Scale (N = 2,222)					*2-Step Scale (N = 2,085)*				
	1	*2*	*3*	*4*	*5*	*1*	*2*	*3*	*4*	*5*
General health	1					1				
Physical ability	.64	1				.58	1			
Energy	.58	.62	1			.50	.54	1		
Resistance to illness	.51	.47	.45	1		.43	.45	.41	1	
Whole life	.40	.33	.36	.30	1	.32	.26	.33	.26	1

Discussion

The findings reveal small differences in frequency distributions between the items administered with the different question forms, and slightly larger effects of question form on relationships among items. On the whole, however, the similarity in results between the two forms is impressive. The data do not support the argument that seven-point scale questions need to be broken down into components for administration on the telephone, at least when the response format involved can be converted to a numerical scale.

Although the findings suggest that the two formats are largely interchangeable, there are a few criteria considering which a preference for the one-step scale can be justified. First, this version produced fewer selections of the most positive category. Since a positive response bias has been noted as a problem with the "satisfaction scale" (Campbell, et al., 1976; Andrews and Withey, 1976), fewer extremely positive mentions may be seen as desirable. Second, the one-step scale produced less missing data (though neither version had much missing data associated with it). Third, the items administered in the one-step form showed higher intercorrelations on the average; this result appears to be due to the typically larger variances associated with items in the two-step version. Finally, interviewers in this study preferred the one-step version, arguing that the two-step form took too long and was repetitious. These complaints are similar to those voiced by the interviewers in Rogers' (1976) study, which involved reading response categories for a series of five-point scales.

As noted above, these conclusions are limited to the situation where the one-step version of a question can be administered as a numerical scale. In questions where nominal response categories are involved or where numerical representation is not possible (e.g. the "delighted"-"terrible" scale), the number of response categories might have to be restricted or a multi-step format employed. The alternative of mailing a visual aid to respondents prior to interviewing may also be efficacious when it is possible to do so.

The conclusions are further limited to the health domain, which may be more salient than other areas. Question format differences might be more likely to appear if respondents are questioned about less salient concerns. Naturally, replication of these results with more items addressing other domains would be highly desirable.

Conclusion

The rapid adoption of the telephone for survey data collection has led to a reexamination of common survey procedures. The research reported here represents one step in the process of testing alternative methods for use in the now common telephone environment. Such investigations are conducted with the knowledge that similar inquiries rarely have been carried out in the personal interview mode, which still serves as the "standard" against which telephone survey findings are judged. Indeed, the logistics and expense of face-to-face interviewing often made this sort of experimentation either difficult or impossible. Thus, when we study alternative methods for telephone surveys, we may provide the only data on such questions in survey methods generally. Much of common survey practice is based on a rich folklore of experience; telephone interviewing offers a means to evaluate more systematically beliefs about which procedures are appropriate and necessary.

Still, the act of evaluation is not simple, even when only two forms of a question are being tested. As can be seen when comparing the Groves and Kahn (1979) investigation and the one reported here, there are a number of important choices to be made in operationalizing the different question formats, labeled "one-step" and "two-step." Groves and Kahn's findings were difficult to interpret, as the form of the questions varied along with the response content. The results of the present study suggest that the response content factor played a major role in their findings. Once the response categories are standardized, as in the present study, it appears that the two question formats produce more comparable results. The lesson to be drawn for future research is that the comparison of question formats must be very carefully specified; one cannot speak generically about "two-step" questions any more than one can equate all "telephone interviews."

References

Andrews, F., and S. Withey
1976 Social Indicators of Well-Being: Americans' Perceptions of Life Quality. New York: Plenus.
Bailey, K. D.
1978 Methods of Social Research. The Free Press.

Campbell, A., P. Converse, and W. Rogers
1976 The Quality of American Life: Perceptions, Evaluations and Satisfactions. New York: Russell Sage Foundation.
Cannell, C., R. Groves, and P. Miller
1981 "The effects of mode of data collection on health survey data." Paper presented at the annual meeting, American Statistical Association, Social Statistics Section.
Durako, S., and T. McKenna
1980 "Collecting health interview survey data by telephone: a mailout experiment." Paper presented at the annual meeting of the American Statistical Association.
Gorden, R. L.
1975 Interviewing: Strategy, Techniques and Tactics (Revised Edition). The Dorsey Press.
Groves, R., and R. Kahn
1979 Surveys by Telephone: A National Comparison with Personal Interviews. New York: Academic Press.
International Educational Resources, Inc.
1976 COFAMM: Confirmatory Factor Analysis with Modifications. Version 1, Release 1.
Kohn, M.
1977 Class and Conformity: A Study in Values (2nd ed.). Chicago: University of Chicago Press.
Moser, C. A., and G. Kalton
1972 Survey Methods in Social Investigation (2nd ed.). New York: Basic Books.
Rogers, T. F.
1976 "Interviews by telephone and in person: quality of responses and field performance." Public Opinion Quarterly 40:51–65.
Schuman, H., and S. Presser
1981 Questions and Answers in Attitude Surveys. Academic Press.

A Comparison of Methods of Increasing Parental Consent Rates in Social Research

TERESA L. THOMPSON

RESEARCHERS have been subject to rules and guidelines governing their use of human participants in research since the concern with human subjects beginning in the late 1960s. One such requirement is the obtainment of informed consent from research participants. When the research participants are children, explicit, written consent frequently must be obtained from the parents or guardian of the child if the study is not directly related to the educational activities of the school.

The requirement for parental consent consistently has resulted in samples limited to half the size of the number that would ordinarily be available (Jessor and Jessor, 1975; Josephson and Rosen, 1978; Kearney, et al., 1983; Lueptow, et al., 1977). While the concern for decreased sample size is an important one in and of itself, some research indicates that limiting a sample to those who are willing to give consent for participation leads to bias in the sample. Singer (1978) found that those who refuse to sign a consent form were also likely to refuse to answer sensitive questions and tended to have either very little or a great deal of education.

Abstract Previous research on children requiring written parental consent has indicated that this requirement reduces samples to half the size they might otherwise be and results in overrepresentation of whites and underrepresentation of blacks. Within the course of a four-year study, four methods were utilized to increase parental consent rates, particularly for black students. All methods increased consent rates. Providing an incentive to the parents increased consent rates significantly more for the white students than for the black students, while communication with the children and incentives to the children resulted in consent rates which were more similar for the two racial groups. Communication with the parents was the most effective method for both black and white children, but was also the most time consuming.

Teresa L. Thompson is a Senior Faculty member, American Center for Management Development, Sarasota, Florida. This research was supported in part by Grant No. G007900083 from the Department of Education and by grants from the University of Delaware. The author thanks Lou Cusella and Dan Slater for helpful comments on an earlier draft of this paper.

Public Opinion Quarterly Vol. 48:779–787 © by the Trustees of Columbia University
Published by Elsevier Science Publishing Co., Inc. 0033-362X/84/0048–779/$2.50

Two studies have investigated how the need to obtain *parental* consent, in particular, will bias a sample. In a comparison of a consenting sample to the entire population, no differences were discovered in grade point average or intelligence scores (Lueptow, et al., 1977). More recently, Kearney, et al. (1983) found that, among whites, parents of students with the highest reading and vocabulary scores were most likely to consent. Those who denied permission had children with somewhat lower scores, and those who did not reply had children with the lowest scores. Among Asian students, the highest scores were observed among those whose parents denied permission, followed by those who consented and those who did not reply. The scores of all three groups of black students were virtually identical. When the sample was compared to the population rather than to those who refused consent, no significant differences were discovered on reading or vocabulary scores. One blatant bias in the sample, however, seemed to be in its racial makeup. Whites were overrepresented, while blacks and Asians were underrepresented. Achieving racial composition in the sample consistent with the population is an ever-present concern to researchers. It appears, then, that what needs to be examined are methods for increasing parental consent rates, particularly among underrepresented groups. Most previous research has used a series of two follow-up letters with consent forms to achieve a 50 percent response rate. Hirschi (1968) had only 12 percent of the parents refuse participation or not reply, but he had field workers visit homes to persuade parents. This method is, in most cases, prohibitively expensive. The goal of the present investigation was to examine other methods of increasing response rates, particularly among black students.

Procedures for Improving Consent Rates

Within the course of a large-scale investigation of the social consequences of the mainstreaming of physically disabled children into "regular" classrooms, parental consent for participation was sought for 1,314 children over the course of four years. Positive responses to the initial request were under 25 percent, so a series of procedures was utilized with different groups of subjects at various times to increase participation. Along with a follow-up letter, the following methods were utilized in different conditions: (1) an incentive to the child for participation; (2) an incentive to the parent; (3) communication with the child about the study; and (4) communication with the parents about the study via a phone call. The overall consent rate and the consent rate for whites vs. blacks were examined.

Methods

SUBJECTS

The participants in the present study included children in grades K–6 in six schools in a countywide school district in a Mid-Atlantic state.[1] The data were collected in the springs of 1979–1982. The school district had begun a busing program in the fall of 1978, so all classes were composed of about 70 percent white students and 30 percent blacks. The nonwhite students in the classes were almost exclusively black; nonwhite students who were not black were not included in analyses because of their small number. All students were of middle or lower middle socioeconomic status.

PROCEDURES

The overall procedures for the study are described elsewhere (Thompson, 1982; 1983). Schools and classes were randomly selected for inclusion in the sample and were randomly assigned to conditions within the constraints of the study of mainstreaming, of which this was just a portion. The parent or guardian of each child in the classes was sent a letter briefly describing the study and a consent form which they were asked to return. The study was conducted during the school day, and individual children were interviewed at some time when the teacher felt it would not interfere with the child's school work. Although only a small percentage of parents refused permission, the very high nonresponse rate initially yielded a sample of 24 percent of the overall group. This percentage was even lower among the black students. Four different methods were used in an attempt to increase response rate. Since funding was limited, the methods were not expensive ones.

Incentive to the child. A portion of the study required taking instant-developing snapshots of all children in some of the classes in order to conduct a sociometric analysis of mainstreamed classrooms. Most of the children appeared eager to have their pictures taken. At the time each child's picture was taken, it was placed on the child's desk so that s/he could watch it develop. The researcher or a research assistant returned to the desk a few moments later, retrieved the picture, and handed the child a letter and consent form to give to his or her parents. The child was told, "If you get your Mom and Dad to sign this form and then bring it back to your teacher, I can give this picture to you at the end of the study."[2] The child, then, was asked to obtain the parent's signature, but not parental consent. The parent checked, "do" or "do not give permission" on the slip. Thus the parent did not have to give consent in order for the child to obtain the picture.

---

[1] The county encompasses the northern third of the state and includes urban, suburban, and rural/farming areas as well as a university. None of the schools sampled, however, were in the university area. The county is the home of major corporations and large automotive plants. County population is approximately 398,000.

Incentive to the parents. The principal investigator of this study had done research on similar topics in the school district in the past. One group of parents received a letter identical to those used as a follow-up in the other conditions but including an offer of copies of articles reporting the results of this previous research. To obtain them, the parent had only to check an appropriate blank on the consent form. The parent did not have to give permission in order to receive the articles, but did have to return the form.

Communication with the child. In some of the classes the researcher, while personally handing the follow-up letter to each child, explained to him or her the basic procedure and goal of the study. While specific hypotheses were not mentioned, the explanation was intended to ease the child's uncertainty about the study and enable him or her to look forward to participation in the study as a "fun experience." Children were told that they would talk to the researcher for about 10 to 15 minutes in the reading room. They were told that they would be shown some pictures of their classmates and asked a couple of questions about them and that they would be shown drawings of other children whom they do not know and asked a couple of questions about what they would say if they were playing with the child in the picture.

Communication with the parent. In some of the classes, parents who had not responded to the first request were called by the researcher on the evening they received the follow-up. The goal and procedures of the study, although not specific hypotheses, were explained. Any questions were answered. All parents could not be reached on the first evening, but 91 percent had been reached in three days.

Each of the four methods resulted in substantial increase in rate of response and consent. The results, categorized by race, are reported in Table 1. Consistent with previous research (e.g., Kearney, et al., 1983; Lueptow, et al., 1977; Singer, 1978) there were no sex differences. Male and female subjects are thus combined in all analyses. As can be seen in Table 1, the numbers in each of the conditions differ. These cell sizes were dictated by other constraints of the study and pragmatic considerations, including time and money available for copying, film, staff help, etc.

---

[2] In actuality, pictures were returned to all children at the conclusion of the study, regardless of whether the form had been returned or not.

Table 1. Response Rates (All Percentages Rounded)

Method	*Overall*	*White*	*Black*	*Prob.*[a]
Follow-up letter	$N = 481$	336	145	
Consent	39%	47%	21%	$p < .05$
No reply	59	51	73	$p < .05$
Deny	2	2	1	$p > .05$
Incentive to child	$N = 450$	315	135	
Consent	80%	82%	79%	$p > .05$
No reply	16	15	20	$p > .05$
Deny	3	3	1	$p > .05$
Incentive to parent	$N = 110$	77	33	
Consent	74%	82%	58%	$p < .05$
No reply	24	16	40	$p < .05$
Deny	2	2	2	$p > .05$
Communication with child	$N = 175$	122	53	
Consent	69%	71%	64%	$p > .05$
No reply	29	27	34	$p > .05$
Deny	2	2	2	$p > .05$
Communication with parent	$N = 98$	68	30	
Consent	95%	97%	91%[b]	$p > .05$
No reply	4	2	8	$p > .05$
Deny	1	1	1	$p > .05$

[a] Based on comparisons of the white and black percentages using the nonparametric test for a difference between two proportions.

[b] This is the percentage of those parents who were reached. A smaller percentage of black parents could be reached by phone than white parents.

Results

COMPARISONS OF THE METHODS

The follow-up letter alone resulted in a consent rate of about 40 percent—higher for whites than for blacks. Here and in all of the other conditions the denial rate was very low. Very few parents who responded were reluctant to give permission.

The incentive to the child resulted in a strong improvement in consent for both white and black children. Both groups of children responded quite similarly—they appeared excited at the possibility of being given a picture of themselves. It is likely that they influenced their parent's interest in the permission form by bringing it to their attention.

The parental incentive also improved responsiveness, but more so among whites than blacks. This method, as used in the present study, does not seem to overcome the problem of differential response/consent rates in the racial groups. Less than half of those responding actually requested that the articles be sent to them. The offer itself increased responsiveness, regardless of whether it was accepted.

Although communication with the child also improved the likelihood of parental response and consent, it did so less dramatically than any of the other methods utilized. This method, however, worked almost as well with the black children as with the white.

The most effective method of increasing parental consent rate was communication with the parents themselves via phone calls. This method resulted in almost unanimous permission from the white parents and very high participation from the black parents. This method, however, was also the most time-consuming.

The consent rates for each of the four methods were compared to those obtained from the use of the follow-up letter alone. This was done for both whites and blacks using the nonparametric test for a difference between two proportions. In all cases, the improvement that was noted was statistically significant ($p<.05$).

Examination of differences at various grade levels indicated only one: the effectiveness of the incentive to the children diminished somewhat with increasing grade, from 91 percent for kindergarteners, to 69 percent for sixth graders. It still, however, improved responsiveness more than any method other than communication with the parents for the black subjects.

As noted earlier, subjects were randomly assigned to conditions within the constraints of the larger study of mainstreaming. One of the dependent measures was a network analysis of the mainstreamed classrooms. This was the measure which required snapshots of the children. As a result, classes could be assigned to the snapshot incentive condition only if they were mainstreamed. A separate analysis was conducted of the mainstreamed classes only, in order to look for effects within groups which were all randomly assigned. The results, which are consistent with those reported above, are presented in Table 2.

Discussion

While some of the methods utilized herein may not be appropriate for some research projects, communication to the children may be of wider applicability. Although in the present study the communication was one-to-one, it may be that explanations from the researcher to the class as a whole would also have some impact. It is likely that the communication functions more on a relational than on a content level (Watzlawick, et al., 1967), since research by Singer (1978) indicated that a written introduction to a survey providing information similar to that given to children here did not affect willingness to respond. Communicating with the researcher, however, may tell the child or even the parent that the researcher seems like a good person who will not hurt me or my child. One-on-one communication is likely to be most effective at accom-

Table 2. Analyses—Mainstreamed Classes Only

Method	*Overall*	*White*	*Black*	*Prob.*[a]
Follow-up letter	$N = 254$	164	90	
Consent	37%	46%	19%	$p < .05$
No Reply	61	54	74	$p < .05$
Deny	2	1	2	$p > .05$
Incentive to child	$N = 450$	315	135	
Consent	80%	82%	79%	$p > .05$
No reply	16	15	20	$p > .05$
Deny	3	3	1	$p > .05$
Incentive to parent	$N = 58$	38	20	
Consent	76%	85%	57%	$p < .05$
No reply	22	13	41	$p < .05$
Deny	2	2	2	$p > .05$
Communication with child	$N = 85$	64	21	
Consent	72%	75%	60%	$p > .05$
No reply	27	23	39	$p > .05$
Deny	1	2	1	$p > .05$
Communication with parent	$N = 50$	35	15	
Consent	96%	98%	90%[b]	$p > .05$
No reply	3	1	9	$p > .05$
Deny	1	1	1	$p > .05$

[a] Based on comparisons of the white and black percentages using the nonparametric test for a difference between two proportions.

[b] Percentage of those parents who were reached.

plishing this, but any communication which results in a positive perception may help.

Researchers should not forget the role of the child in the parental consent process. The two child-oriented methods used herein resulted in closer response rates in the two racial groups than is typically reported (e.g., Kearney, et al., 1983). Children may have a strong impact on their parents' interest in the permission request simply by increasing their awareness of it.

Involving children in the parental consent process, of course, raises some ethical concerns. In addressing these concerns, we must first assume that the research project is an important one with a goal of providing information which will be of assistance to parents and children. In the present study, the purpose was the improvement of social opportunities for physically disabled children and the assessment of the effectiveness of mainstreaming as a vehicle for the accomplishment of this. The resultant information, then, has social utility. Further, the reward provided to the children was not one which could in any way bring harm to them, as candy, toys, or money potentially could. Since it was a picture of themselves, it was more of an intrinsic motivator, functioning, for example, as a self-concept builder through the provision of self-information (Cusella, 1984; Deci, 1975). The reward involved no lying to children, as praise when used as a reward might, and did not even cause discomfort to them if the parent did not sign because *all* children were given their pictures at the conclusion of the study.

Finally, the children themselves were not making the decision. A responsible adult, the parent or guardian, still maintained decision-making status. Motivating the child simply increased the chances that the slip would arrive at home and come to the parent's attention, rather than being left at school, on the bus, or on the playground. Follow-up phone calls to 45 (10 percent) of the parents participating in an incentive-to-the-child condition indicated no cases of unusual pressure for consent being applied to the parent by the child or examples of situations where parents had ostensibly received forms but not actually signed them.

The sizes of some of the cells were small enough to limit the generalizability of some of the present findings. In particular, the percentage of black parents who could be reached by phone was less than that of white parents, and resulted in a fairly small group.

Two other methods of increasing responsiveness were eliminated during pretesting because of their ineffectiveness, but might be successfully varied for other researchers. Explanations to the classes by teachers were eliminated because most teachers reported that they felt uncomfortable and therefore neglected to include details of the study in their explanations. Also, meetings with groups of parents were scheduled for evenings in order to provide them with information but, when only one parent attended the first meeting and none the second, the meetings were discontinued.

Further research should look for differences in parental consent rates for different types of topics. The present topic—mainstreaming and attitudes toward the disabled—was only moderately sensitive. Results may well be different for topics of greater sensitivity.

References

Cusella, L.P.
1984 "The effects of feedback source, message and receiver characteristics on intrinsic motivation." Communication Quarterly 32:211–21.
Deci, E.L.
1975 Intrinsic Motivation. New York: Plenum Press.
Hirschi, T.
1969 Causes of Delinquency. Los Angeles: University of California Press.
Jessor, R., and S.L. Jessor
1975 "Adolescent development and the onset of drinking." Journal of Studies on Alcohol 36:27–51.
Jessor, R., S.L. Jessor, and J. Finney
1973 "A social psychology of marijuana use: longitudinal studies of high school and college youth." Journal of Personality and Social Psychology 26:1–15.

Josephson, E., and M. Rosen
1978 "Panel loss in a high school drug study." Pp. 115–32 in D. B. Kandel (ed.), Longitudinal Research on Drug Use. New York: Wiley.
Kearney, K.A., R.H. Hopkins, A.L. Mauss, and R.A. Weisheit
1983 "Sample bias resulting from a requirement for written parental consent." Public Opinion Quarterly 47:96–102.
Lueptow, L., S.A. Mueller, R.R. Hammes, and L.R. Master
1977 "The impact of informed consent regulations on response rate and response bias." Sociological Methods and Research 6:183–204.
Singer, E.
1978 "Informed consent: consequences for rate and response quality on social surveys." American Sociological Review 43:144–62.
Thompson, T.L.
1982 "The development of listener-adapted communication in physically handicapped children: a cross-sectional study." Western Journal of Speech Communication 46: 32–44.
1983 "Communication with the handicapped: a three year study of the effectiveness of mainstreaming." Communication Education 32:185–95.
Watzlawick, P., J. Beavin, and D.D. Jackson
1967 The Pragmatics of Human Communication. New York: W.W. Norton & Co.

The Content of Survey Introductions and the Provision of Informed Consent

JEFFERY SOBAL

WHAT DO survey researchers tell potential respondents about themselves and their research in the introductions to their surveys? Introductions are a key component of surveys. Yet they have received relatively less attention than other aspects of the survey process which may be less important, such as the type of postage stamp used in mail surveys, the dress of personal interviewers, or the average number of rings needed in telephone interviews. Most norms about what to include in survey introductions are set by precedent, not by experiment. New survey researchers are socialized into the existing "folklore" about the appropriate contents of survey introductions by example, not evidence.

Recommendations about survey introductions in the literature on interviewing emphasize establishing the legitimacy of the interviewer, eliciting cooperation of the potential respondent, and enhancing rapport in the interview (Cannell and Kahn, 1968; Dobey, 1967; Gordon, 1969; Kahn and Cannell, 1957; Parten, 1966; Richardson, et al., 1965; Selltiz, et al., 1976; Survey Research Center, 1969). Other methodologically oriented considerations of interview introductions have focused upon the consequences of different introductions. When introductions disclosed more information, refusals to participate in surveys were found to be more frequent by Blumburg, et al. (1974), and Kearney, et al. (1983); less frequent by Hauck and Cox (1974); and about the same by Josephson and

Abstract Information disclosed to potential respondents in survey introductions was examined in a sample of 78 introductions obtained by mail from AAPOR members. The following information was disclosed: research organization, in 86 percent of the introductions; interviewer, 82 percent; research topic, 81 percent; sponsor, 45 percent; confidentiality, 42 percent; survey purpose, 26 percent; future data use, 24 percent; anonymity, 24 percent; sampling technique, 21 percent; survey length, 13 percent; statement that participation was voluntary, 10 percent; sample size, 4 percent; and consent signature requirement, 4 percent. Mail surveys (and to a lesser extent those by academic institutions and on sensitive topics) provided more introductory information. For ethical reasons, such as providing informed consent, as well as methodological reasons, professional norms about providing information in survey introductions need to be examined.

Jeffery Sobal is Assistant Professor and Associate Director of Research, Department of Family Medicine, University of Maryland School of Medicine.

Public Opinion Quarterly Vol. 48:788–793 © by the Trustees of Columbia University
Published by Elsevier Science Publishing Co., Inc. 0033-362X/84/0048–788/$2.50

Rosen (1978), Leuptow, et al. (1977), Singer (1978a), and Sobal (1982). Data quality and bias have also been examined in relation to the amount of information disclosed in survey introductions. Less bias was reported in surveys with fuller introductions by Cannell, et al. (1977), and minimal differences were found by Singer (1978a, 1978b, 1979, 1982), Singer and Khonke-Aguirre (1979), and Sobal (1982). Although the traditional advice is that introductions should be brief and disclose minimal information about the research, most experimental studies of introductions suggest that fuller disclosure of information incurs few if any methodological costs, and some benefits.

In contrast to the utilitarian approach to disclosing information in interview introductions taken by methodologists, an ethical perspective on the issue is based upon moral principles such as autonomy (Bower and DeGasparis, 1978; Kelman, 1968, 1982). Advocates of this viewpoint emphasize the potential respondents' right to be provided the fullest possible information about the survey before deciding whether or not to give their "informed consent" to participate.

There is at present no consensus about what constitutes "informed" consent, and "informed" is considered to be a continuous rather than a categorical state (Wolfensberg, 1967). While informed consent may be encouraged by codes of ethics (Bower and DeGasparis, 1978) or required by sponsoring agencies (U.S. DHEW, 1971), the norms of the profession are the major determinant of how much information is actually provided. At this time, we do not know how much information is usually given in survey introductions, or how this varies by type of organization or type of survey. This study provides a tentative answer by describing the contents of a sample of introductions from surveys done by members of the American Association of Public Opinion Research (AAPOR).

Methods

A questionnaire was sent to 172 AAPOR members randomly sampled from the 1980 membership directory. They were asked to send a copy of the introduction to their most recent survey and to answer several questions about the study, including: type of survey (personal interview, telephone, mail), sensitivity of questions on the survey (somewhat, not at all), and the organization conducting the survey (academic, business, other). A self-addressed, stamped envelope was included. When the introductions were received, they were coded as disclosing or not disclosing the following types of information: interviewer's name, study director's name, organization conducting the research, organization sponsoring the research, purpose of the study, promise of confidentiality, promise of anonymity, sampling procedure, sample size, subject of the study, length of time the survey will take, future use of the data, a statement that participation was voluntary, and request to sign an informed consent form.

Results

Of the 172 AAPOR members contacted, 108 responded, for a 63 percent response rate with no follow-ups. Of the 108 replies, only 78 (45 percent) were usable because some declined to participate for various reasons and others returned the questionnaire without including a survey introduction or vice versa. There was no consistent pattern of response by academic, business, or other organizations.

The percentage of introductions disclosing the different types of information varied considerably, as shown in Table 1. Over 80 percent of the introductions disclosed research organization, interviewer (or study director for mail surveys), and research topic. The sponsor of the survey was disclosed 45 percent of the time, significantly more often in mail surveys than in telephone or personal interviews ($p < .05$).

Confidentiality was mentioned in 42 percent of the introductions, and was included significantly more often in mail than in personal interviews or telephone surveys ($p < .05$), and more often by academic than other organizations ($p < .01$). Anonymity was mentioned in 26 percent of the introductions. Only 50 percent of the introductions which mentioned confidentiality also included a statement of anonymity, but 80 percent of the 20 introductions which promised anonymity also included a statement about confidentiality. Thus, 47 percent of the introductions promised either confidentiality or anonymity.

Table 1. Types of Information Disclosed In Survey Introductions

Type of Disclosure	*%*	*(n)*
Research organization	85.9%	(67)
Interviewer[a]	82.1	(64)
Research topic	80.8	(63)
Sponsor	44.9	(35)
Confidentiality	42.3	(33)
Anonymity	25.6	(20)
Purpose	25.6	(20)
Future data use	24.4	(19)
Sampling technique	20.5	(16)
Survey length	12.8	(10)
Participation voluntary	10.3	(8)
Sample size	3.8	(3)
Consent signature	3.8	(3)

[a] Includes study director's name mentioned in mail surveys.

The purpose of conducting the survey was disclosed in 26 percent of the introductions, significantly more often in mail surveys than telephone or personal interviews ($p < .01$). The 21 percent of introductions which mentioned sampling technique were more frequently from academic than business or other organizations ($p < .05$), and more frequently for surveys including sensitive or threatening questions ($p < .05$). Only a small percentage of the survey introductions provided information about survey length, the voluntary nature of participation in the survey, sample size, or a requirement for a consent signature.

Discussion

The data reported here show that survey introductions are not homogeneous, suggesting that varied norms about the content of introductions exist among survey researchers. Mail surveys disclose more than telephone or personal interviews, perhaps because no interviewer is available to elaborate upon the information provided in the introduction. Academic organizations disclose a few more types of information, possibly because of different norms in these settings, or constraints in the academic environment, such as Institutional Review Boards (IRB's). The sensitivity or threat of questions appears to have little effect on the information disclosed to potential respondents.

Because survey research depends upon the good will of the public, current practices with respect to survey introductions need reassessment. This is especially true in light of the issue of respondent burden (Bradburn, 1978; Frankel and Sharp, 1981; Sharp and Frankel, 1983) and suggestions that the public is becoming reluctant to enter the role of respondent (Frankel, 1976; Petry, 1976). Survey researchers need to devote more attention to the content of introductions. Standards about survey introductions that take both methodological and ethical considerations into account should be developed and disseminated.

References

Blumberg, Herbert H., Carolyn Fuller, and A. Paul Hare
1974 "Response rates in postal surveys." Public Opinion Quarterly 38:113–23.
Bower, Robert T., and Priscilla DeGasparis
1978 Ethics in Social Research: Protecting the Interests of Human Subjects. New York: Praeger.
Bradburn, N.
1978 "Respondent burden." Pp. 49–43 in Health Survey Research Methods, DHEW Publication No. (PHS) 79-3207.
Cannell, C.F., and R.L. Kahn
1968 "Interviewing," In E. Aronson (ed.), Handbook of Social Psychology, Vol. 2, 2nd ed. Reading, MA: Addison Wesley.

Cannell, C.F., L. Oksenberg, and J.M. Converse
1977 "Striving for response accuracy: experiments in new interviewing techniques." Journal of Marketing Research 14:306–15.
Dobey, John T. (ed.)
1967 An Introduction to Social Research. New York: Appleton-Century-Crofts.
Frankel, J., and L.M. Sharp
1981 "Measurement of respondent burden: summary of study design and early findings" Statistical Reporter, No. 81–4:105–12.
Frankel, Lester
1976 "Statisticians and people—the statistician's responsibility." Journal of the American Statistical Association 71:9–16.
Gordon, Raymond L.
1969 Interviewing: Strategy, Techniques and Tactics. Homewood, Ill: Dorsey
Hauck, M., and M. Cox
1974 "Locating a sample by random digit dialing." Public Opinion Quarterly 38:253–56.
Josephson, E., and M.A. Rosen
1978 "Panel loss in a high school drug study." In D.B. Kandel (ed.), Longitudinal Research on Drug Use: Empirical Findings and Methodological Issues. New York: Wiley.
Kahn, Robert L., and Charles F. Cannell
1957 The Dynamics of Interviewing: Theory, Techniques and Cases. New York: Wiley.
Kearney, K.A., R.A. Hopkins, A.L. Mauss, and R.A. Weisheit
1983 "Sample bias resulting from a requirement for written parental consent." Public Opinion Quarterly 47:96–102.
Kelman, Herbert C.
1968 A Time to Speak: On Human Values in Social Research. San Francisco: Jossey-Bass.
1982 "Ethical issues in different social science methods." Pp 40–100 in T.L. Beauchamp, et al. (eds.), Ethical Issues in Social Science Research. Baltimore: Johns Hopkins University Press.
Leuptow, L., S.A. Mueller, R.R. Hammes, and L.R. Master
1977 "The impact of informed consent regulations on response rate response bias." Sociological Methods and Research 6:183–204.
Parten, Mildred
1966 Surveys, Polls and Samples: Practical Procedures. New York: Cooper Square Publishers.
Petry, Glenn H.
1976 "Questionnaire on questionnaires finds: too many questionnaires." Phi Delta Kappan 57(5):350.
Richardson, Stephen A., Barbara S. Dohrenwend, and David Klein
1965 Interviewing: Its Form and Functions. New York: Basic Books.
Selltiz, C., L. Wrightsman, and S.W. Cook
1976 Research Methods in Social Relations, 3rd ed. New York: Holt, Rinehart and Winston.
Sharp, L.M., and Frankel, J.
1983 "Respondent burden: a test of some common assumptions." Public Opinion Quarterly 47:36–53.
Singer, Eleanor
1978a "Informed consent: consequences for response rate and response quality in social surveys." American Sociological Review 43(2):144–61.
1978b "The effect of informed consent procedure on respondent's reactions to survey." Journal of Consumer Research 5:49–57.
1979 "Informed consent procedures in surveys: some reasons for minimal effects upon response." In M.L. Wax and J. Cassell (eds.), Federal Regulations: Ethical Issues and Social Research. Boulder, CO: Westview Press.
1982 "Informed consent procedures in telephone interviews." American Sociological Review 47:416–27.

Singer, Eleanor, and Luanne Khonke-Aguirre
1979 "Interviewer expectation effects: a replication and extension." Public Opinion Quarterly 43(2):245–60.

Sobal, Jeffery
1982 "Disclosing information in interview introductions: methodological consequences of informed consent." Sociology and Social Research 66(3):348–61.

Survey Research Center
1969 Interviewers Manual. Ann Arbor, Michigan: Institute for Social Research.

U.S. DHEW
1971 The Institutional Guide to DHEW Policy on Protection of Human Subjects (DHEW Publication No. NIH 72-102). Washington, D.C.: U.S. Government Printing Office.

Wolfensberg, W.
1967 "Ethical issues in research with human subjects." Science 155:47–51.

Incentives for Increasing Return Rates: Magnitude Levels, Response Bias, amd Format

J. SCOTT MIZES, E. LOUIS FLEECE, AND CINDY ROOS

THE use of monetary and nonmonetary incentives has received significant attention as a method of increasing return rates from mail surveys. Armstrong (1975), in a review of eight studies, concluded that monetary incentives have a substantial impact on survey return rate, typically reducing nonreponse by a third. Nonmonetary incentives also have positive effects, although the response increase has not always been statistically significant (cf. Nederhof, 1983). Initial evidence from direct comparisons suggests that monetary incentives are superior to nonmonetary ones in increasing mail survey returns (Goodstadt, et al., 1977).

Armstrong's review concluded that increasing the size of a monetary incentive increases mail survey response. However, he notes that few studies have compared different magnitudes of monetary incentives, and few studies have investigated the effects of incentives greater than a dollar, leaving the effects of larger incentives unknown. Thus, the current study was designed to address these issues by comparing the effects of $5 and $1 incentives on mail return rate.

Abstract The use of monetary incentives has been shown to significantly increase response rate. However, previous investigations have rarely investigated the effects of incentives greater than $1, compared different magnitudes of incentives, or investigated response bias due to incentives. The current study also investigated the utility of an Answer Check. Results suggest that monetary incentives increase response rate, larger incentives do not necessarily further increase survey response, incentives do not appear to bias responses, and the Answer Check does not facilitate response rate.

J. Scott Mizes is Assistant Professor, Department of Psychology, North Dakota State University, Fargo, ND 58105. E. Louis Fleece is Assistant Professor of Psychiatry (Psychology), University of Alabama—Birmingham School of Medicine, and Psychologist, Birmingham VA Medical Center. Cindy Roos is a graduate student, Department of Psychology, University of Alabama—Birmingham. This research was supported by a grant to J. Scott Mizes from the Department of Psychiatry, University of Alabama—Birmingham School of Medicine.

Public Opinion Quarterly Vol. 48:794–800 © by the Trustees of Columbia University
Published by Elsevier Science Publishing Co., Inc. 0033-362X/84/0048-794/$2.50

Nederhof (1983) has noted that there have been few investigations of the effect of incentives on quality of responses received. Incentives may result in a self-selection process or demand artifact which may alter subjects' responses (Sawyer, 1975). Nederhof (1983) found no difference in the responses to fixed answer questions as a function of a nonmonetary incentive. Goodstadt et al. (1977) found no difference in mean attitude ratings toward a magazine for monetary and no incentive conditions. Hansen (1980) found no significant difference in the information derived under monetary and nonmonetary incentive conditions. These are precious few studies on this question. Thus, the current study compares responses received under monetary and no incentive conditions to assess the presence of response bias.

The final focus of this study is to assess the effects of an Answer Check. The Answer Check has the survey answer blanks printed on the back of the check, and thus the questionnaire responses return through the Federal Reserve System as canceled checks. The rationale is to link the incentive and desired behavior as closely as possible. Enclosure of a check and a separate answer sheet allows the respondent to cash the check and then answer the questionnaire at a later time thereby increasing the probability of forgetting to complete the questionnaire, losing the questionnaire, etc. The Answer Check forces respondents to answer the questionnaire before they cash the check rather than afterwards. Once the Answer Check is cashed, the answer sheet is gone and the opportunity to respond is lost.

The following hypotheses guided the investigation: (1) enclosure of a monetary reinforcer would result in increased return rate, (2) a larger monetary reinforcer would result in a higher return rate, and (3) the Answer Check would result in higher return rate than simple enclosure of a check and a separate answer sheet. No prediction was made on the effect of monetary incentives on response quality.

Method

The current experiment was conducted March through June 1982 as part of a study on rhinitis medicamentosa, i.e., chronic intranasal inflammation due to excessive use of nasal spray decongestants. Five-item questionnaires on the incidence, treatment, and success of treatment of rhinitis medicamentosa were mailed to 200 physicians specializing in allergy. The sample was randomly selected from the larger pool of physicians listed under Allergy or Allergy/Immunology in the telephone directory yellow pages of major metropolitan areas across the United States.

Each questionnaire packet contained a cover letter on university letterhead stationery which detailed a definition of rhinitis medicamentosa and a brief description of the purpose of the research. While the questionnaires were photocopied, some personalization was present. A personal salutation and handwritten closing signature appeared on each letter; however, a personal inside address was not used. Also enclosed was the questionnaire and (depending on condition) an answer sheet printed on either a self-addressed postcard or on a bank check.

There were 40 subjects in each of five conditions. In the control condition, potential respondents were mailed a cover letter, questionnaire, and answer postcard. In the \$1/Postcard and \$5/Postcard conditions, a cover letter, questionnaire, answer postcard, and bank check were mailed. An additional paragraph on the standard cover letter indicated that the enclosed check was a small gratuity for the respondents' time and trouble. In the \$1/Answer Check and \$5/Answer Check conditions, a cover letter, questionnaire, and bank check were mailed. The answer blanks were printed on the back of this check. An additional paragraph on the standard cover letter instructed respondents to answer on the back of the check, and said that the questionnaire would return to us as a cancelled check, and that the check was a gratuity for their effort.

Results

Data were converted from a frequency to a dichotomous score format and analyzed via ANOVA procedures, as recommended by Linton and Gallo (1975). Nine questionnaires were "returned to sender," and therefore the replacement of missing data for unequal subjects procedure was utilized, as described by Linton and Gallo (1975). Pairwise comparisons of all experimental groups to the control group were conducted via planned comparisons utilizing the Least Significant Difference Test. Subjects in the control condition returned 52.6 percent of the questionnaires; \$1/Answer Check, 52.6 percent; \$5/Answer Check, 71.8 percent; \$1/Postcard, 73.7 percent; \$5/Postcard, 73.7 percent. Planned pairwise comparisons revealed that the \$1/Postcard and \$5/Postcard conditions differed significantly from control ($p \leq .05$, LSD = 20.7 percent), a trend toward a significant difference was found for the \$5/Answer Check condition ($p \leq .10$, LSD = 19.2 percent), and no difference was found for the \$1/Answer Check condition, $p > .05$. The experimental conditions were further examined via a 2(Money) × 2(Postcard vs. Answer Check) ANOVA. The Money, $F(1,149) = 1.70$, Postcard versus Answer Check $F(1,149) = 2.44$, and interaction factors $F(1,149) = 1.71$ were all nonsignificant, $p > .05$.

Subjects' responses to three questions on frequency of rhinitis medicamentosa patients and percentage successfully treated were compared across groups in three one-way ANOVAS. Responses did not differ significantly for any of the three questions $F(4,105) = 1.142$, $F(4,118) = 1.35$, and $F(4,118) = .33$, all $p > .05$. Means and standard deviations for questions and conditions are presented in Table 1. Inspection of the table shows no consistent trend of different data being obtained depending on experimental condition. Thus, the statistically nonsignificant findings do not appear to be due to small sample size, but rather accurately reflect the findings.

Discussion

The $5 and $1 monetary incentive increased return rate by 21.1 percent, while the $5 answer check resulted in a 19.2 percent increase. While the latter finding was marginally statistically significant, all three findings represent marked increases from a practical standpoint. Greater statistical significance might have been achieved had a sample size more characteristic of survey studies (600–1,000) been used. Nonetheless, the increase in return rate in this study compares well with previous studies on the effect of monetary incentives.

Results of the two-way ANOVA suggest that $5 has no more positive effect than $1 return rate. This finding is contrary to the observation by Armstrong (1975), based on studies using incentives of $1 or less, that larger incentives lead to increased response rate. Armstrong attempted to specify more precisely the relationship between response rate and incentive size with the "rule of thumb" that larger incentives lead to

Table 1. Question Responses for Each Incentive/Answer Check Condition

		Condition			
Question	*Control*	*$1/ Postcard*	*$5/ Postcard*	*$1/ Answer Check*	*$5/ Answer Check*
% Rhinitis medicamentosa patients					
Mean	4.3	7.5	7.3	7.4	7.1
Standard deviation	3.6	14.0	7.7	14.8	7.3
Total rhinitis medicamentosa patients					
Mean	15.3	37.3	17.2	30.2	21.3
Standard deviation	22.6	67.2	15.0	58.6	23.0
% Patients withdrawn from nasal spray					
Mean	92.9	83.3	79.1	78.0	82.0
Standard deviation	9.6	24.1	23.5	30.6	25.3

reductions in *nonresponse* up to a maximum of 40 percent. Interestingly, the \$1 and \$5 incentives reduced nonresponse in this study by approximately 40 percent. Thus, \$1 may have achieved maximal reduction in nonresponse, making \$5 unnecessary. These findings would suggest that a given sample could be divided into three response probability groups: those who respond readily, those who will respond readily if prompted, and those who are very unlikely to respond. This latter group may be relatively impervious to incentives or other response inducements unless the incentives are at exceptionally high levels. Thus, monetary incentives may increase mail survey response up to a point (the 40 percent reduction in nonresponse suggested by Armstrong?), further increases may have little additional beneficial effect, and very large increases in incentives may finally elicit response from a recalcitrant remainder of the sample.

With reference to the effect of incentives with physicians, the findings of Gunn and Rhodes (1981) are interesting to contrast with the current findings. These researchers compared physicians' actual completion rate of structured personal interviews under no-incentive, \$25, and \$50 incentive conditions. While caution regarding the comparability of findings must be exercised because of the potential differences attributable to a mailed questionnaire versus personal interviews, the similarity of response rate is notable. They found a 58 percent response rate for no incentive, 69 percent for \$25, and 77 percent for \$50 incentive conditions. The difference between \$25 and \$50 was nearly statistically significant, $p < .09$.

Depending on one's viewpoint, Gunn and Rhodes' findings could be viewed as either consistent or inconsistent with ours. One could argue that their small sample size ($N = 600$) precluded finding a significant but "real" effect at traditional .05 significance levels. However, the failure to achieve a clearly significant effect with what "should be" a strong effect is notable. Factors to consider here are that their sample was much larger than ours, and their incentive size and *difference* between incentive levels was large. Also, Heberlein and Baumgartner (1981), in a meta-analysis review of the questionnaire response rate literature, found that incentives tended to increase response rate, but that the correlation between incentives and response rate was not significant. This argues against a simple linear trend for incentive effects, and would be more consistent with Armstrong's and our model.

It is also interesting to note that the maximal response (77 percent) in the Gunn and Rhodes (1981) study represented a reduction in nonresponse of 45 percent, notably close to Armstrong's (1975) suggestion of a maximal reduction of nonresponse of 40 percent. Even if one accepts Gunn and Rhodes' \$25 versus \$50 incentive level difference as real, it could be argued that the very high \$50 incentive was enough to entice the participation of the "recalcitrant few," according to the model we suggested earlier. Thus, while the size of the incentive may need to be larger for personal interviews of physicians (due to greater cost for the respondent), what may still be applicable is the basic model of increasing incentives facilitating improved response up to a point of diminishing returns, with further increases coming with only very large incentives. Clearly, research with *many* incentive levels, covering small to very large incentives, is needed to address this question.

The current findings indicate no systematic bias of responses due to monetary incentives. These findings are consistent with three previous studies addressing this question. This accumulation of evidence is encouraging since the use of incentives is so attractive from a response rate vantage point. However, caution should be exercised before concluding definitely that incentives do not induce response bias. It is possible that researchers have yet to examine response bias under conditions or on dependent variable where differences would be found.

No evidence was found supporting the use of the Answer Check over a regular check with separate answer sheet, as evidenced by the lack of a significant effect of this factor on the two-way ANOVA. In fact, the Answer Check may have a negative effect. The $1 Answer Check resulted in the same return rate as did the control condition. Given that a plain $1 check did significantly increase return rate, one could conclude that the Answer Check detracted from the return rate. However, this did not occur for the $5 Answer Check. A possible explanation is that because respondents are unfamiliar with the Answer Check, it lacks credibility. While $1 may not be sufficient to overcome this deleterious effect, $5 may be. It should be noted that the different results of the two-way and one-way ANOVA are conceptually confusing, but probably represent a statistical aberration due to different error terms and degrees of freedom.

The generality of the current findings may be suspect because of the short questionnaire length and the specialized population. Heberlein and Baumgartner (1981) did not find a significant correlation between response rate and any of several measures of questionnaire length. Only when several response-facilitating variables were controlled did length show a relationship to response rate. However, this relationship was found only for length as determined by number of questions (versus number of pages, for example) and the relationship was very small ($r = -.05$). Thus, the current questionnaire's length is not likely to be a serious impediment to generality of the findings. One hypothesis is that the greatest impact of questionnaire length may be on incentive magnitude, not on the hypothesized nonlinear relationship between incentive size and response rate.

Given the Gunn and Rhodes (1981) finding of different response rates even for different physician specialities, and Heberlein and Baumgartner's (1981) finding of significant population effects, the limits on generality of the current findings due to the specialized population is apparent. This is particularly true for generalizing to the general population. However, the hypothesis again is that the population sampled may have more impact on the incentive size needed, rather than on the linear or nonlinear relationship between incentive size and response rate.

In summary, the following conclusions can be drawn from the current study: (1) monetary incentives significantly increase response rate; (2) larger incentives do not *necessarily* lead to greater increases in response rate, i.e., there is a point of diminishing returns, and furthermore, this point may be quite far away from 100 percent return rate; (3) incentives appear to bias responses little; and (4) an Answer Check does not facilitate, and may harm, response rate.

References

Armstrong, J. Scott
1975 "Monetary incentives in mail surveys." Public Opinion Quarterly 39:111–16.
Goodstadt, Michael S., Linda Chung, Reena Kranitz, and Gaynoll Cook
1977 "Mail survey response rates: their manipulation and impact." Journal of Marketing Research 14:391–95.
Gunn, Walter J., and Isabelle N. Rhodes
1981 "Physician response rates to a telephone survey: effects of monetary incentive level." Public Opinion Quarterly 45:109–15.
Hansen, Robert A.
1980 "A self-perception interpretation of the effect of monetary and nonmonetary incentives on mail survey respondent behavior." Journal of Marketing Research 17:77–83.
Heberlein, Thomas A., and Robert Baumgartner
1978 "Factors affecting response rates to mailed questionnaires: a quantitative analysis of the published literature." American Sociological Review 43:447–62.
Kerlinger, F. N.
1973 Foundations of Behavioral Research (2nd ed.). New York: Holt, Rinehart, and Winston.
Linton, Marigold, and Phillip S. Gallo, Jr.
1975 The Practical Statistician: Simplified Handbook of Statistics. Monterey, CA: Brooks/Cole.
Nederhof, Anton J.
1983 "The effects of mail incentives: two studies." Public Opinion Quarterly 47:103–11.
Sawyer, Alan G.
1975 "Demand artifacts in laboratory experiments in consumer research." Journal of Consumer Research 4:20–30.

Bias in a Directory Sample for a Mail Survey of Rural Households

FREDERICK J. KVIZ

A MAJOR concern for mail survey design is obtaining a representative and accurate list of the population under study, which often is difficult for surveys of the general population of relatively large geographic areas. Telephone directories often are adopted as the sampling frame for general population mail surveys because they are readily available at little or no cost. However, while 93 percent of all private households in the United States are estimated to have telephone service, according to the 1980 U.S. Census, telephone directories have been shown to be biased against representation of lower socioeconomic status households, which disproportionately do not have telephones or have unpublished telephone numbers (Perry, Jr., 1968; Brunner and Brunner, 1971; Leuthold and Scheele, 1971; Roslow and Roslow, 1972; Glasser and Metzger, 1975). Although exclusion bias is serious in most large metropolitan areas, it may be rather small in rural areas. Therefore, telephone directories often appear useful as sampling frames for general population surveys of rural areas.

Abstract This study examines bias in a mail survey of rural households that used telephone directories as the sampling frame and compares it with a parallel survey that used an independent sample selected from field listings. A higher response rate was obtained with the directory sample, which was primarily attributed to poorer address information for the listed sample. Although coverage of the study population was higher for the listed frame than for the directory frame, no significant differences were found for eight respondent and seven household characteristics. Also, no significant differences were found for seven key content variables regarding health behaviors and attitudes. The findings support the use of telephone directories as the sampling frame for general population mail surveys of rural areas where such coverage of the population is considered adequate.

Frederick J. Kviz is an Associate Professor in the Epidemiology-Biometry Program, School of Public Health, and Research Associate Professor in the Survey Research Laboratory at the University of Illinois at Chicago. This research was supported by Grant RO 1 HS 02778 from the National Center for Health Services Research. Sampling, data collection, and data processing were conducted by the Survey Research Laboratory at the University of Illinois. The author is grateful to Johnny Blair, Linda B. Lannom, and Jutta P. Sebestik for assistance in tabulating some of the results.

Public Opinion Quarterly Vol. 48:801–806 © by the Trustees of Columbia University
Published by Elsevier Science Publishing Co., Inc. 0033-362X/84/0048–801/$2.50

However, an evaluation of bias in mail surveys using directory samples seems never to have been reported. It would be useful to examine the size and nature of bias in such surveys and to consider whether that bias might be substantially reduced. To that end, this paper examines bias in a general population mail survey of rural households using telephone directories as the sampling frame in comparison with a mail survey using an independent sample selected from field listings of the same population. The listed sampling frame was expected to provide more complete coverage of the study population because it would avoid exclusion bias of directories due to nonsubscribers to the telephone service and unpublished telephone numbers.

Method

The universe for the study consisted of households in counties identified as at least 50 percent rural according to the 1970 U.S. Census in Illinois (i.e., at least one-half of the households were located in places with a population less than 2,500). Because the principal objective of the project from which the data are reported was to examine utilization of and perceived need for health services, the universe was divided into five strata according to values on the Index of Medical Underservice (see Federal Register, 1975). One county was selected within each stratum with probabilities proportionate to the number of housing units per county. One half of the sample in each county was randomly selected from the most recent telephone directories. The other half of the sample was selected independently, using a multistage area probability design, from field listings obtained for face-to-face interviews that were conducted in the same counties as part of another phase of the larger study. Field listing was performed during February and March, 1978.

According to the 1980 U.S. Census, 91 percent of the households in the five counties were estimated to have a telephone available, and approximately 95 percent of rural telephone subscribers were estimated to be listed in the directories (Sudman, 1976). Therefore, the percentage of the study population included in the directory sampling frame is estimated to be 86 percent (91% × .95).

A listed sampling frame is rarely, if ever, perfectly complete and accurate, because of demolitions and new construction in the period between the original listing and data collection, and because of households that are missed during listing or that are listed incorrectly. No data are generally available regarding the accuracy of field listing. However, considering the short time interval between the completion of field listing and data collection, and using unpublished data for a similar survey recently conducted by the Survey Research Laboratory at the University of Illinois, we estimated exclusion bias for the listed sample to be about one percent.

Data were collected during late spring and early summer of 1978. The mail survey design for both the directory and listed samples was a modification of the design described by Dillman, et al. (1974) and consisted of an initial mailing of the questionnaire by first-class mail, a follow-up letter mailed one week later, and two follow-up mailings of the questionnaire at approximately four-week intervals. The third mailing of the questionnaire was sent via certified mail and was preceded by a telephone reminder call to all nonrespondents for whom a telephone number could be obtained. Finally, telephone interviews were attempted with all nonrespondents following the third mailing of the questionnaire.

For the directory sample, questionnaires were addressed to the household member in whose name the telephone was listed. For the listed sample, questionnaires were addressed to the household member whose name was obtained (from a mailbox or personal inquiry) during listing. The instructions indicated that the questionnaire could be completed by any adult (18 years or older) member of the household who was knowledgeable about the household's recent health history and current health practices.

Findings

SURVEY OUTCOMES

Response rates for the two sampling frames were compared to assess potential nonresponse bias differences. As displayed in Table 1, the highest overall response rate was obtained with the directory sample. Although nonresponse bias is the term usually applied to the complement of the response rate, this difference was not due to greater cooperation by respondents in the directory sample, as indicated by the similar cooperation rates. The main reason for this difference was that address information was poorer for the listed sample, as evidenced by a post office return rate for the listed sample that was more than twice that for the directory sample.

While field listing generated adequate descriptions of household location for use by interviewers (nonresponse bias for the face-to-face interview phase of the study with a parallel version of the questionnaire administered to an independent sample in the same counties was only 14.2 percent), it was difficult to obtain useful mailing addresses for many of the rural households. Also, because telephone numbers could not be obtained for all households in the listed sample that did not return a completed mail questionnaire, relatively fewer telephone interviews were

Table 1. Outcomes for Directory and Listed Samples for a Mail Survey of Rural Households

Outcome[a]	*Directory (n = 389)*	*Listed (n = 371)*	*Significance*[b]
Response rate			
Mail	53.4	47.4	ns[c]
Telephone	6.8	4.6	ns
Total	60.3[d]	52.0	$p < .05$
Cooperation rate	65.9	64.0	ns
Completion rate	56.6	51.8	ns
Post office return rate	8.5	18.7	$p < .001$
Eligibility rate	93.8	99.5	$p < .001$

[a] Rates were computed as follows:

(1) *Response rate* = percent of eligible sample members from whom a completed questionnaire was collected.

(2) *Cooperation rate* = percent of eligible sample members contacted from whom a completed questionnaire was collected.

(3) *Completion rate* = percent of entire sample from whom a completed questionnaire was collected.

(4) *Post office return rate* = percent of initial questionnaires mailed to eligible sample members returned as not deliverable.

(5) *Eligibility rate* = percent of entire sample eligible.

[b] Two-sided test for differences between proportions.

[c] Not significant.

[d] Sum of mail and telephone follow-up response rates does not equal total response rate because of rounding.

completed with nonrespondents for the listed sample than for the directory sample. Although the difference in eligibility rates was statistically significant, with the listed sample being more successful in identifying households located within the study counties, this difference was not very substantial.

RESPONDENT AND HOUSEHOLD CHARACTERISTICS

The extent to which differential response for the two sampling frames was related to basic sociodemographic variables was examined. No statistically significant differences were found for any of eight respondent characteristics: sex, age, education, marital status, employment status, relationship to head of household, race, and length of residence at address. Similarly, no statistically significant differences were found for any of seven respondent household characteristics: per capita income, residence type, residence ownership, household size, children age 17 or younger, adults age 18 to 64, and adults age 65 or older.

HEALTH BEHAVIORS AND ATTITUDES

The main concern for a study of survey bias is whether differences are observed in responses to the main variables in the analysis. The study from which the data for this paper are taken focused on seven variables that measured four dimensions of primary health care services assessment: (1) access and availability (have a usual source of care); (2) use of services located outside county of residence during previous 12 months; (3) utilization (number of nonemergency, outpatient visits to a physician during the previous 12 months); (4) health status (number of serious symptoms, from a list of 11, experienced during previous 12 months; (5) visit to a physician regarding any illness or injury during previous 12 months; (6) perceived health status); (7) general satisfaction with available health care services. A comparison of responses for the two sampling frames revealed no statistically significant differences for any of these seven key variables.

Discussion

Although coverage of the study population was higher for the listed than for the directory sample, the two samples did not differ significantly for 15 respondent and household characteristics. Therefore, the study population appears to have been equally well represented by either sample. This conclusion is supported by the finding that the two samples also did not differ significantly for seven key content variables regarding health behaviors and attitudes.

The directory sample was significantly more efficient in terms of a better response rate. It must be noted, however, that the response rate for both samples was rather low despite intensive follow-up efforts. Although it cannot be stated conclusively, this may be due to the considerable length of the questionnaire, which consisted of 28 pages and required approximately 40 minutes to administer in a face-to-face interview. At any rate, despite the low response rates, both samples appear to have adequately represented the population of households. This conclusion is based on a comparison of household characteristics for the mail surveys and the previously mentioned face-to-face interview survey that was conducted in the same counties. The only statistically significant difference was that the number of adults age 65 or older was slightly higher for the mail surveys. This difference is probably due to a reluctance on the part of the elderly to admit a strange person (interviewer) into their home because of a concern for their personal safety.

The purpose of any sample survey is to collect data that are sufficiently representative of the study population so that inferences may be made confidently from the sample about the population. Because researchers do not have access to unlimited funds, survey designs must also be efficient. In this instance, the findings strongly support the conclusion that telephone directories may provide an acceptable and efficient sampling frame for general population mail surveys of rural areas. Although this study was conducted in only one state, the findings may be generalizable to rural populations in other states in the Northeast and North Central (Census) regions of the United States, where at least 98 and 92 percent, respectively, of the households are reported in the 1980 Census as having a telephone. Caution must be exercised in extending these findings to some states in the South and West, where the percentage of households with a telephone is as low as 82 and 86 percent, respectively, and supplementation of directory listings might be advisable.

References

Brunner, James A., and G. Allen Brunner
1971 "Are voluntarily unlisted telephone subscribers really different?" Journal of Marketing Research 8:121–24.
Dillman, Don A., James A. Christenson, Edwin H. Carpenter, and Ralph M. Brooks
1974 "Increasing mail questionnaire response: a four-state comparison." American Sociological Review 39:744–56.
Federal Register
1975 "Designation of medically underserved areas and population groups." Vol. 40 (September 2):40315–320.
Glasser, Gerald J., and Gale D. Metzger
1975 "National estimates of nonlisted telephone households and their characteristics." Journal of Marketing Research 12:359–61.
Leuthold, David A., and Raymond Scheele
1971 "Patterns of bias in samples based on telephone directories." Public Opinion Quarterly 35:249–57.
Perry, Joseph B., Jr.
1968 "A note on the use of telephone directories as a sample source." Public Opinion Quarterly 32:691–95.
Roslow, Sydney, and Laurence Roslow
1972 "Unlisted phone subscribers are different," Journal of Advertising Research 12:35–38.
Sudman, Seymour
1976 Applied Sampling. New York: Academic Press.

Comments and Letters

On Beatty and Walter

If a social scientist, because of some unusual personal qualities or other resources, is able to penetrate the higher and inner circles of the upper social class, establish rapport with a delinquent street gang, or do extended naturalistic observation with a tribe in some remote jungle rainforest, the observations and data analysis are virtually beyond the scrutiny and checking by any other social scientist. At the other extreme, analysts of public opinion data who report new analyses of archival surveys that are widely available, using standard and widely accepted, that is, nonesoteric, statistical procedures, are completely open to checking by colleagues. In fact, one valuable teaching exercise in a modern public opinion course is to have students replicate, as nearly and completely as possible, the analyses in an article using such a data set. But this vulnerability to committing *recognizable* errors in analyses is one of the significant hazards of doing secondary analysis.

Recently, Beatty and Walter ("Religious preference and practice: reevaluating their impact on political tolerance," POQ 48: 318–29) reported an interesting and potentially worthwhile analysis of the link between religion and tolerance of social deviants, using the NORC surveys of 1976, 1977, and 1980. Unfortunately, their analysis began with an error in constructing the index pertaining to Communists. As a consequence, some of the numbers in each of their six tables are incorrect, and the article is thereby flawed.

In their appendix, they give the three tolerance questions, and then state, "The *same* three questions were asked for: A person claiming that blacks are inferior, a Communist, a person who advocates doing away with elections and letting the military run the country, and a homosexual" (p. 328, emphasis added). Here they inadvertently omitted the Atheist from their list of deviants, but that was not the crucial error. In fact, the same questions were *not* asked. The free speech and library book questions were the same for all five deviant categories, but the teaching college question was the same for only four. In the case of the Communist, the question was reversed and the meaning of the code was thereby reversed. For the other deviant categories, the question was, "Should

Public Opinion Quarterly Vol. 48:809–819 © by the Trustees of Columbia University
Published by Elsevier Science Publishing Co., Inc. 0033-362X/84/0048–809/$2.50

such a person be allowed to teach in a college or university, or not?" with no (coded 5) being the *intolerant* alternative and yes (coded 4) being the *tolerant* alternative. However, in the case of the Communist, the question was, "Suppose he is teaching in a college. Should he be fired, or not?" So in this case, yes (coded 4) is the *intolerant* answer, while no (coded 5) is the *tolerant* answer. (This is admittedly a strange pattern, but I assume it was done this way to facilitate comparisons with baseline data from the 1950s.)

Thus, if one erroneously assumes that the same three questions are asked in regard to each of the five social deviant categories and proceeds on that basis, as Beatty and Walter seem to have done, to build a comparable intolerance index for each deviant category, a person gets the highest score possible on the intolerance of Communist index by giving one tolerant and two intolerant answers.

What difference does this make? The table below shows a correlation matrix for the five tolerance indices, giving the published correlations of Beatty and Walter (from their Table 6) above and to the right of the diagonal, and the correlations, as I think they ought to be, below and to the left of the diagonal.

Correlations (*r*) Among Five Tolerance Indices

	Atheists	*Communists*	*Homosexuals*	*Racists*	*Militarists*
Atheists	X	.49	.64	.64	.63
Communists	.71	X	.48	.38	.44
Homosexuals	.65	.66	X	.50	.58
Racists	.66	.56	.52	X	.55
Militarists	.65	.67	.59	.57	X

Beatty and Walter report results that imply the improbable conclusion that attitudes toward Communists and Atheists are more *weakly* linked (.49) than are attitudes toward Homosexuals and Atheists (.64), Racists and Atheists (.64), or Militarists and Atheists (.63). My analysis suggests that, on the contrary, the link between attitudes toward Communists and Atheists is the *strongest* of all 10 correlations. (Incidentally, the strong anti-Communist feelings within the U.S. are probably attributable, in part, to the link between Communism and Atheism.)

From the preceding, it follows that column 2 in Tables 1 and 2 in Beatty and Walter's article is incorrect. Furthermore, Beatty and Walter derive for each person a "least liked group," that is, the deviant category toward which the person is most intolerant. For that score to have utility or meaning, however, the indices for each deviant category must be correct. That not being the case, their results for the "least liked group" (column 7 in Table 1, column 6 in Table 2, Tables 4 and 5) are undoubtedly wrong since the wrong group would get designated as "least liked" for many people. The average tolerance scores (Table 3 and column 6 in Table 1) would also be off somewhat in that the averages are based, in part, on the flawed intolerance of Communists index.

I have not argued that the broad substantive conclusions drawn by Beatty and Walter are incorrect, although in some cases they might be. Some more specific implications from their article are obviously in error. For instance, their results imply that Lutherans are more tolerant of Communists than they are of Atheists or Racists, when, in fact, they are more tolerant of Atheists or Racists than they are of Communists. Overall, their initial error in constructing the Communist index causes so many subsequent errors that, as it stands, the article is next to worthless. Beatty and Walter are undoubtedly correct in alleging a strong relationship between religiosity and intolerance of social deviants, but that is not exactly a novel thesis, and unfortunately, in light of this comment, it cannot be said that their article advances our understanding of the topic. For there to be a contribution in their article, all of their calculations would need to be repeated, and all of their tables would need to be corrected.

What lesson is to be drawn from this? Obviously, for people doing secondary analysis, there is no substitute for proceeding cautiously and reading the codebooks with great care. It might also be a good idea to have one's analyses subjected to an independent check prior to publication.

DONALD GRANBERG
University of Missouri

Response to Granberg

Professor Donald Granberg has correctly pointed out a mistake we made in the coding of a variable. We are obviously very embarrassed by the mistake and apologize to the Editor and the readers of this journal. Fortunately, this mistake was on only one of the three variables in the Communist index, and the Communist index is one of five in the main dependent variables; the Correlation between the corrected variable and the incorrect one is .92.

One would expect that if the thesis of the paper is valid a correction of the coding error would strengthen relationships. Professor Granberg has pointed out that this is in fact the case in Table 6 and it is also true elsewhere. In Tables 1 and 2 the Eta for religion and the Communist index increases from .24 to .28 and the Eta between religion and least liked index increases to .29. When using attendance as the independent variable, the Eta between it and the Communist index increases from .13

to .19. The least liked relationship does not change. In Table 4 the F increases from 6.62 to 11.45 and the Beta increases slightly to .17. In Table 5 the F increases to 47.03 and the beta increases to .15.

The important question is whether or not our results are invalid after the correction. First, we wished to demonstrate that there were important differences among Protestant denominations. Table 1 clearly demonstrates this to be the case and the argument is strengthened with the correction. The second conclusion was that frequent attendance is related to intolerance; and here again, the conclusion is not modified by the correction.

The third conclusion stated that these relationships would remain following controls for socioeconomic status. This conclusion is also valid after correction. Finally, we wished to demonstrate that although the object of intolerance may be important, as Sullivan has pointed out, intolerant individuals tend to be intolerant regardless of the object of intolerance. This finding is even more clearly apparent following the Granberg corrections to Table 6.

For the convenience of readers, corrected columns for the original tables are shown on pages 813–14.

Professor Granberg should be complimented for finding this coding mistake. Because of his discovery, the relationships in the article have been strengthened.

B. OLIVER WALTER
University of Wyoming

Corrections for Table 1

Denomination	Communists	X	Least Liked
None	1.65	1.58	2.11
Congregationalist	1.70	1.81	2.24
Jewish	1.80	1.79	2.46
Episcopalian	1.88	1.86	2.58
Presbyterian	2.27	2.11	2.77
Mormon	2.34	2.13	2.87
Jehovah's Witness	2.12	1.93	2.93
United Church of Christ	2.28	2.23	2.97
Catholic	2.33	2.18	2.97
Lutheran	2.34	2.45	2.97
Methodist	2.57	2.43	3.10
Other Protestant	2.55	2.42	3.13
Christian	2.49	2.42	3.13
Church of Christ	2.87	2.75	3.25
Seventh Day Adventist	2.60	2.26	3.00
Baptist	2.77	2.62	3.33
Assembly of God	3.03	2.89	3.38
Pentecostal	2.94	2.68	3.49
Church of God	3.24	3.02	3.63
F =	18.03		20.7
Eta =	.28		.29

Corrections for Table 2

Church Attendance	Communists	Least Liked
Never	2.12	2.72
At least once a year	2.07	2.67
Once a year	2.22	2.78
Several times a year	2.37	2.95
Once a month	2.35	3.01
Two or three times a month	2.49	3.07
Nearly every week	2.62	3.22
Every week	2.58	3.15
More than once a week	2.87	3.37
F =	19.4	19.0
Eta =	.19	.18

Corrections for Table 3[a]

Denomination	Gamma[b]	Mean Difference	Low Attenders	High Attenders
Episcopalian	.06	.04	2.43	2.47
Presbyterian	.03	.14	2.81	2.95
Catholic	.10	.24	2.77	3.01
Lutheran	.12	.28	2.80	3.08
Methodist	.09	.23	3.02	3.25
Other Prot.	.38	.59	2.80	3.39
Baptist	.23	.41	3.12	3.53

[a] The F = test for the effect of denomination on tolerance, controlling for attendance is 18.7. The F for the effect of attendance, controlling for denomination is 64.8. The overall F is 25.7.

[b] Gamma tests the relationship between the frequency of attendance and political tolerance for each denomination.

Corrections for Table 4[a]

Religious Group	*Unadjusted Mean*	*Mean Adjusted for Controls*
None	2.09	2.44
Congregational	2.20	2.49
Jewish	2.39	2.77
Episcopalian	2.59	2.91
Presbyterian	2.75	2.89
Mormon	2.85	2.99
Jehovah's Witness	2.88	2.64
United Church of Christ	2.94	2.83
Catholic	2.94	2.89
Lutheran	2.96	2.98
Methodist	3.07	3.13
Other Protestant	3.12	3.10
Christian	3.17	3.11
Church of Christ	3.19	3.06
Seventh Day Adventist	3.00	2.83
Baptist	3.29	3.12
Assembly of God	3.33	3.07
Pentecostal	3.59	3.32
Church of God	3.63	3.49

F for religion controlling for attendance, education, and occupational prestige = 11.45 $p = .001$. Beta = .17.

[a]When the same method of analysis is used controlling for father's occupation and education when respondent was 16 years of age, the results were as follows:

F for religion, controlling attendance, father's education and occupational prestige = 7.67. $p = .001$. Beta = .21.

The measure of occupational prestige used was that developed by Hodge, et al.

Corrections for Table 5[a]

Frequency of Attendance	*Unadjusted Mean*	*Mean Adjusted for Controls*
Once a year of less	2.70	2.76
Several times a year	2.98	2.95
Nearly every week or more	3.20	3.16

F for attendance controlling for religion, education and occupation prestige = 47.03. Significant to .001. Beta = .15.

[a]When the same method of analysis is used controlling for father's occupation and education when respondent was 16 years of age, the results were as follows:

F for religion, controlling attendance, father's education and occupational prestige = 49.3. p = .001. Beta = .12.

The Polls:
Britain at the Polls 1945–1983

ROBERT M. WORCESTER

Public Opinion is everything.—ABRAHAM LINCOLN

You ask me, Sir, whether I lead or follow public opinion?
The truth is I meet it on the way. —EDMUND BURKE

THE LATE great George Gallup replied to criticism that repeated polls themselves have undue influence on voters on voting day by saying, "One might as well insist that a thermometer makes the weather." The evidence is in Britain that public opinion polls do influence the way some people vote and that they're being taken into account by increasingly large numbers of people, especially in by-election situations where tactical voting is more prevalent. After the last general election 4.5 percent of British voters said they were influenced by the polls.[1] This was twice as true of Liberal/SDP Alliance supporters as for supporters of the more established Conservative and Labour parties.

In 1983 the Labour Party lost the election as no party in Britain has lost an election before. The combination of a too left-wing manifesto (platform), combined with a divided Labour Party, combined with a leader (Michael Foot) thought to be inadequate for the job resulted in a 10-point lower Labour share of the vote from the lowest figure recorded before; 16 percentage points lower than the average of their share in the 11 general elections since 1945. The Conservatives, on the other hand, in 1983 received 44 percent share, almost exactly their 11-election average.

As Table 1 shows, the changing force has been the rise in the Liberal share in the two 1974 and 1979 elections and the establishment of Britain's fourth party of Social Democrats and their alliance with the Liberals which, in the closing days of the 1983 General Election, threatened to overtake the Labour Party as the alternative party to the Government. Indeed, if you look at a colored map of

---

[1] Gallup Poll, BBC, June 9–10, 1983, [illegible] c.4,000

Robert M. Worcester, an American, is Chairman and Managing Director of Market & Opinion Research International, Ltd. (MORI), and has for 15 years served as public opinion polling advisor to the Labour Party in Great Britain. He is retained by The (London) Times, Sunday Times, and Economist newspapers.

Public Opinion Quarterly Vol. 48:824–833 © by the Trustees of Columbia University
Published by Elsevier Science Publishing Co., Inc. 0033-362X/84/0048–824/$2.50

Table 1. Voting in General Elections in Great Britain: 1945–1983

Year	*Conservative* %	*Labour* %	*Liberal* %	*Conservative Lead* +%	*Labour Lead* +%
1945	40	49	9		+ 9
1950	43	47	9		+ 4
1951	48	49	3		+ 1
1955	49	47	3	+ 2	
1959	49	45	6	+ 4	
1964	43	45	11		+ 2
1966	41	49	9		+ 8
1970	46	44	8	+ 2	
1974 February	39	38	20	+ 1	
1974 October	37	40	19		+ 3
1979	45	38	14	+ 7	
			SDP/Liberal Alliance		
1983	44	28	26	+16	

Great Britain, the Labour strongholds in the major conurbations, South Wales, north of England and Scotland, it shows how in what isolated pockets Labour's strength resides. But an even more impressive geographical pattern is shown by the number of constituencies, covering vast areas, in which Labour fell to third place.

Over the past decade, younger people in Britain have grown to be over one-third of the electorate, having been just over a quarter in 1974. But more important to the electoral situation has been the rise in those people objectively assessed as middle class, from 35 percent to 40 percent, and the compensating reduction in the working class, from 65 percent to 60 percent of the electorate. Further, in Britain as in America, trade unionists have over the decade declined as a percentage of the electorate.

As a result of the Conservative Party policy to sell off local authority houses to sitting tenants at a discount, there has been a substantial increase in owner-occupier houses. In fact, for several elections now housing tenure has been the single best determinant factor of voting intention. Among those working-class voters who own their own home, 25 percent more voted Tory than Labour in 1983; those working-class electors living in Council houses voted 21 points Labour. Thus, between 1979's general election and June 1983 the structural change in the profile of British voters in every instance of sex, class, trade union membership, area of country, and housing tenure favored the Conservatives (Table 2).

A more detailed examination of the voting behavior of the British public is shown in Table 3, which indicates for the key demographic classifications of sex, age, class, and trade union membership the voting behavior in the four general elections in the past decade. Also, using the aggregate of the fortnightly tracking polls conducted by MORI in Britain, Table 3 shows the public's voting intention since the last election. The most remarkable thing about the Tories' victory in 1983 was its homogeneity. Overall, the protest vote of defecting Labour supporters went to the Alliance, which more than doubled the 1979 Liberal share. In the electorate as a whole, the Tories actually declined a point between the two general elections. Labour lost 10 points (as noted above) and other parties

Table 2. Changes in the Profiles of British Voters, 1974–1983

	1974	*1979*	*1983*	*Change 1979–83*
	%	%	%	±%
Total	100	100	100	
Sex				
Male	48	51	48	–3
Female	52	49	52	+3
Age				
18–24	11	12	13	+1
25–34	16	21	21	0
35–54	35	31	32	+1
55+	38	36	34	–2
Class				
ABC1 (Middle-class: white collar and professional)	35	35	40	+5
C2 (Skilled working-class)	32	33	31	–2
DE (Unskilled working-class)	33	32	29	–3
Trade unionist				
Yes	28	30	25	–5
No	72	70	75	+5
Area				
North (inc. Scotland)		37	35	–2
Midlands (inc. Wales)		24	24	0
South		39	41	+2
Housing				
Owner-occupier	50	52	58	+6
Council (local authority)	53	35	29	–6

SOURCE: ORC: '74 MORI: '79–'83.

(mostly represented by the Scottish National Party in Scotland and Plaid Cymru in Wales) also declined one point. This rise and fall was, within a point or two, uniform by sex, age, class, and trade union membership.

Following the disaster that was the General Election of 1983, the Labour Party held remarkably steady throughout the summer. In Britain, Party Conferences are held annually and the major event of the Party Conference season of the autumn of 1983 was the race for the leadership of the Labour Party between Welsh MP Neil Kinnock, the eventual victor, and Birmingham MP Roy Hattersley. The effect of the changeover of leadership from the discredited Michael Foot to Neil Kinnock was marked sharply in the polls. From a 28 percent share at both the General Election and immediately pre-Conference, the Labour Party's share reached nearly to its February 1974 level in the immediate post-Conference MORI Poll.

Swing,[2] a particularly useful British concept, was 6 percent to the Labour Party as soon as the new leader had been elected. The 9-point rise came 6 points from the Alliance, 2 points from the Tories, and 1 point from other parties. The Conservatives' 16-point lead was cut to a mere 5 points. In subsequent quarters

---

[2] The difference between the two leading parties divided by two, over two elections or polls.

Table 3. Voting Intention Aggregate Analysis: 1983–1984

	General Election Results				Net Change General Election from 1979 to GE 1983	Aggregate MORI Results				
	Feb 74	Oct 74	May 79	June 83		Pre-conference (IIIQ '83)	Post-conference (IVQ '83)	Jan–March 1984 (IQ '84)	Apr–June 1984 (IIQ '84)	Change (IQ '84–IIQ '84)
	%	%	%	%						
All (Actual Result)	(*n* = 3,900)	(3,858)	(5,672)	(13,926)		(6,474)	(7,649)	(8,236)	(9,780)	
Con	39	37	45	44	−1	45	42	41	40	−1
Lab	38	40	38	28	−10	28	37	39	38	−1
Lib (Alliance)	20	19	14	26	+12	26	20	18	20	+2
Other	3	4	3	2	−1	1	1	2	2	
Con lead	+1	−3	+7	+16		+17	+5	+2	+2	
Swing to Conservative		−2	+5	+4.5		+0.5	−6	−7	−7	
Men	(*n* = 1,873)	(1,795)	(2,889)	(6,710)		(3,082)	(3,677)	(3,453)	(4,702)	
Con	35	32	43	42	−1	42	40	39	39	0
Lab	42	43	40	30	−10	31	40	41	40	−1
Lib (Alliance)	19	18	13	25	+12	26	19	18	19	+1
Other	4	7	4	2	−2	1	1	2	2	0
Con lead	−7	−11	+3	+12		+11	+0	−2	−1	
Swing to Conservative		−2	+7	+4.5		−0.5	−5.5	−7	−6.5	
Women	(*n* = 2,023)	(2,063)	(2,783)	(7,166)		(3,392)	(3,971)	(4,283)	(5,078)	
Con	42	39	47	46	−1	47	43	43	41	−2
Lab	35	38	35	26	−9	26	35	37	36	−1

Table 3. (Continued)

	General Election Results				Net Change General Election from 1979 to GE 1983	Aggregate MORI Results				
	Feb 74	Oct 74	May 79	June 83		Pre-conference (IIIQ '83)	Post-conference (IVQ '83)	Jan–March 1984 (IQ '84)	Apr–June 1984 (IIQ '84)	Change (IQ '84–IIQ '84)
	%	%	%	%						
Lib (Alliance	21	20	15	27	+12	26	21	18	22	+4
Other	2	3	3	2	−1	1	1	2	1	−1
Con lead	+7	+1	+12	+20		+21	+8	+5	+5	
Swing to Conservative		−3	+5.5	+4		+0.5	−6.5	−7.5	−7.5	
18–24	(n = 394)	(377)	(697)	(1,8[illegible])		(972)	(1,141)	(1,182)	(1,495)	
Con	24	24	42	42	0	42	41	38	36	−2
Lab	46	42	41	33	−8	33	38	43	44	+1
Lib (Alliance)	25	27	12	23	+11	23	19	16	17	+1
Other	5	7	5	2	−3	2	2	3	3	0
Con lead	−22	−18	+1	+9		+9	+3	−5	−8	
Swing to Conservative		−2	+8.5	+4		0	−3	−7	−8.5	
25–34	(n = 688)	(780)	(1,193)	(2,825)		(1,239)	(1,512)	(1,605)	(1,854)	
Con	35	33	43	40	−3	43	41	40	38	−2
Lab	40	38	38	29	−9	29	36	38	37	−1
Lib (Alliance)	21	24	15	29	+14	26	21	20	24	+4
Other	4	5	4	2	−2	2	2	2	1	−1
Con lead	−5	5	+5	+11		+14	+5	+2	+1	
Swing to Conservative		0	+5	+3		+1.5	−3	−4.5	−5	

35–54	(*n* = 1,412)	(1,486)	(1,764)	(4,412)		(2,175)	(2,498)	(2,793)	(3,244)	
Con	37	34	46	44	−2	44	42	43	41	−2
Lab	40	42	35	27	−8	27	36	36	36	0
Lib (Alliance)	19	20	16	27	+11	27	21	20	22	+2
Other	4	4	3	2	−1	2	1	1	1	0
Con lead	−3	−8	+11	+17		+13	+6	+7	+5	
Swing to Conservative		−2.5	+9.5	+3		−2	−5.5	−5	−6	
55+	(*n* = 1,402)	(1,239)	(2,018)	(4,769)		(2,088)	(2,495)	(2,656)	(3,187)	
Con	46	42	47	47	0	47	42	43	42	−1
Lab	33	40	38	27	−11	27	39	39	38	−1
Lib (Alliance)	19	14	13	24	+11	25	18	17	19	+2
Other	2	4	2	2	0	1	1	1	1	0
Con lead	+13	+2	+9	+20		+20	+3	+4	+4	
Swing to Conservative		−5.5	+3.5	+5.5		0	−8.5	−8	−8	
ABC1	(*n* = 1,387)	(1,548)	(1,967)	(5,542)		(2,648)	(3,160)	(3,375)	(3,851)	
Con	57	56	59	55	−4	57	57	54	51	−3
Lab	20	19	24	16	−8	15	22	23	23	−1
Lib (Alliance)	21	21	15	28	+13	26	22	21	24	+3
Other	2	4	2	1	−1	2	1	1	2	+1
Con lead	+37	+37	+35	+39		+42	+35	+30	+28	
Swing to Conservative		0	−1	+2		+1.5	−2	−4.5	−5.5	
C2	(*n* = 1,299)	(1,292)	(1,907)	(4,276)		(1,978)	(2,339)	(2,487)	(2,926)	
Con	29	26	41	40	−1	39	38	37	36	−1
Lab	48	49	41	32	−9	34	42	44	43	−1
Lib (Alliance)	19	20	15	26	+11	26	19	18	19	+1
Other	4	5	3	2	−1	1	1	1	2	+1
Con lead	−19	−23	0	+8		+5	−4	−7	−7	
Swing to Conservative		−2	+11.5	+4		−1.5	−6	−7.5	−7.5	

(Continued)

Table 3. **(Continued)**

	General Election Results				*Net Change General Election from 1979 to GE 1983*	*Aggregate MORI Results*				
	Feb 74	*Oct 74*	*May 79*	*June 83*		*Pre-conference (IIIQ '83)*	*Post-conference (IVQ '83)*	*Jan–March 1984 (IQ '84)*	*Apr–June 1984 (IIQ '84)*	*Change (IQ '84–IIQ '84)*
	%	%	%	%						
DE	(*n* = 1,210)	(1,126)	(1,798)	(4,060)		(1,848)	(2,151)	(2,374)	(3,003)	
Con	28	22	34	33	−1	33	28	29	28	−1
Lab	49	57	49	41	−8	41	53	53	53	0
Lib (Alliance)	19	16	13	24	+11	25	18	16	17	+1
Other	4	5	4	2	−2	4	1	2	2	0
Con lead	−21	−35	−15	−8		−8	−25	−24	−25	
Swing to Conservative		−7	+10	+3.5		0	−8.5	−8	−8.5	
Trade unionist	(*n* = 1,053)	(1,059)	(1,715)	(3,495)		(1,648)	(1,920)	(2,097)	(2,503)	
Con	23	23	33	31	−2	32	31	30	29	−1
Lab	55	55	51	39	−12	37	47	49	48	−1
Lib (Alliance)	19	16	13	29	+16	29	21	19	22	+3
Other	3	6	3	1	−2	2	1	2	1	−1
Con lead	−32	−32	−18	−8		−5	−16	−19	−19	
Swing to conservative		0	−7	+5		−1.5	−4	−6.5	−6.5	

SOURCE: ORC: '74 MORI: '79–'83

that lead has been further cut to only 2 points, as the Tories' share has continued to tail off and the Alliance has added back a point or two. The Alliance recovery has principally been among women in the 25–34 age group and among the middle class.

The 1983 general election, if a disaster for the Labour Party, was a triumph for the polls. The four eve-of-poll polls that were conducted face-to-face—Gallup, Marplan, NOP, and MORI—had an average error per party share of 0.8 percent, i.e., within ±1 percent. Even including two telephone polls, by Harris (a recall of their earlier face-to-face survey) and Audience Selection, who were by a long way the furthest out, the average error was 1.20, i.e., well within the ±2 percent figure. In fact, the average error per party in the 1979 election was 0.95, but then no one was polling on the telephone then (Tables 4 and 5).

Table 4. Predictions of the Polls: General Election 1979

Q. How would you vote if there were a General Election tomorrow? ASK THOSE UNDECIDED/REFUSED
Q. Which party are you most inclined to support?

Poll	*Gallup*	*Marplan*	*NOP*	*MORI*	*Actual Result*
Newspaper	Telegraph	Sun	Mail	Ev. Std	3rd May
Sample size	2,348	1,973	1,069	1,089	
Fieldwork	30–1 May	1 May	1–2 May	2 May	
Method	Quota	Quota	Quota	Quota	
	%	%	%	%	
Conservative	43.0	45.0	46.0	45.0	44.9
Labour	41.0	38.5	39.0	37.0	37.7
Liberal	13.5	13.5	12.5	15.0	14.1
Other	2.5	3.0	2.5	3.0	3.3
Conservative lead	2.0	6.5	7.0	8.0	7.2
Error on gap	−5.2	−0.7	−0.2	+0.8	
Average error	1.65	0.45	1.2	0.5	0.95

Table 5. Predictions of the Polls: General Election 1983[a]

Poll	*Audience Selection*	*Harris*	*Gallup*	*Marplan*	*NOP*	*MORI*	*Actual Result*
Newspaper	Sun	Observer	Telegraph	Guardian	N. Cliffe	Ev. Std	9 June
Sample size	1,100	576	2,003	1,335	1,013	1,101	
Method	(Tel)	(Tel recall)	Quota	Quota	Quota	Quota	
	%	%	%	%	%	%	%
Conservative	45	47	45.5	46	47	44	44
Labour	23	25	26.5	26	25	28	28
Liberal	29	26	26	26	26	26	26
Other	3	2	2	2	2	2	2
Conservative lead	22	22	19	20	22	16	16
Error on gap	+6	+6	+3	+4	+6	0	
Average error	2.25	1.5	0.75	1.0	1.5	0	1.2

[a] Same questions as Table 4

All major polling organizations in Great Britain have now switched over to quota samples, whereas several had been known for random sampling in the past. Virtually everyone in Britain is automatically enrolled annually in a relatively effective registration system, yet the empirical results over the past decade have convinced all major pollsters that the speed of turnaround of quota samples, more or less tightly controlled, is preferred to the random methodology which requires callbacks.

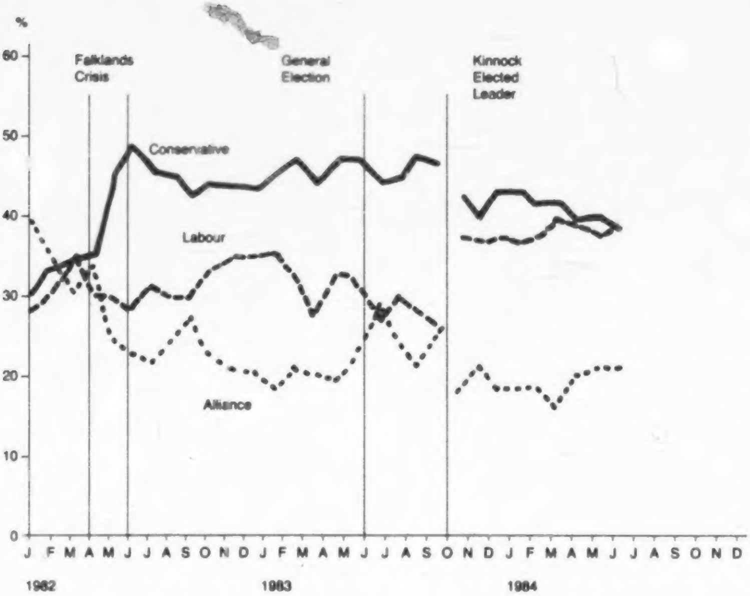
Of 43 polls that have been published on or around election days in Britain, in the 12 general elections that have been conducted since the war, the average error per party has been 1.5 percent, the largest error in 1951 and 1970 being 2.6 percent and the smallest, in 1979, less than 1 percent (Table 6).

In graphic terms, the shift in voting intention, including the "Falklands Factor," is shown in Figure 1. There are several other things that are demonstrated by this graph: (1) that the Alliance was on the slide before the Falklands War began, (2) that the Falklands Factor didn't wear out, (3) that the Labour Party suffered badly at the hands of the Alliance during the few weeks of the 1983 British General Election and (4) that the discontinuity in voting intention with the election of Kinnock as leader of the Labour Party was remarkable, and (5) that the Labour Party lives again. At the last poll before publication the Conservatives and Labour parties were "level pegging" at 39 percent each with the Alliance trailing badly behind at 20 percent.

A similar pattern is shown in Figure 2. When respondents were asked, "Are you satisfied or dissatisfied with the way the Government is running the country?" Mrs. Thatcher's performance as Prime Minister consistently ran ahead of that of the satisfaction level of the Government, and the election of Kinnock as leader of the Labour Party had a quick rise in the first half of 1984 but a quick falloff as well. Mrs. Thatcher's satisfaction level has been dropping steadily for nearly a year and it may well be that before she calls the next general election in 1987 or 1988 (a privilege of the British Prime Minister saving the loss of a vote of confidence in the House of Commons or the expiration of a five-year term), Mrs. Thatcher may well again touch the point she reached in 1980 at the end of her first year as Prime Minister as "the least popular Prime Minister in British history."

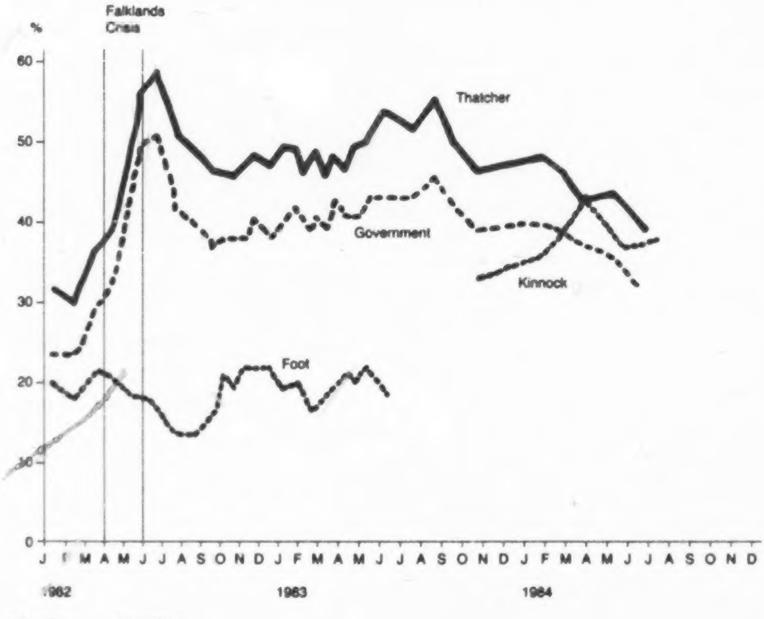
Table 6. Accuracy of the Polls, 1945–1983

Year	*Average Error per Party*	*Number of Polls*
1945	1.6%	1
1950	1.5	2
1951	2.6	3
1955	0.7	2
1959	1.1	3
1964	1.3	4
1966	1.7	4
1970	2.6	5
1974 February	1.6	5
1974 October	1.3	4
1979	0.9	4
1983	1.2	6
Average	1.5	43



SOURCE: MORI.

Figure 1. Voting Trends 1982–1984



SOURCE: MORI.

Figure 2. Satisfaction Trends 1982–1984

Q. Are you satisfied or dissatisfied with the way:
The government is running the country?
Mrs. Thatcher is doing her job as Prime Minister?
Mr. Foot is doing his job as Leader of the Opposition?

A0610-1

POPQAE 49(1) 1–142 (1985)
ISSN 0033–362X

Public Opinion Quarterly

Early Calls on Election Results and Exit Polls: Pros, Cons, and Constitutional Considerations
J. Ronald Milavsky, Al Swift, Burns W. Roper, Richard Salant, and Floyd Abrams

Uncovering Rape: The Watchdog Press and the Limits of Agenda Setting
David L. Protess, Donna R. Leff, Stephen C. Brooks, and Margaret T. Gordon

Television News, Real-World Cues, and Changes in the Public Agenda
Roy L. Behr and Shanto Iyengar

A Critical Analysis of Australian Television Coverage of Election Opinion Polls
Ted J. Smith III and Derek O. Verrall

The Failure of Liberal/Conservative Ideology as a Cognitive Structure
Norman R. Luttbeg and Michael M. Gant

Does Time of Voting Affect Exit Poll Results?
Ronald J. Busch and Joel A. Lieske

The Importance of Replicating a Failure to Replicate: Order Effects on Abortion Items
George F. Bishop, Robert W. Oldendick, and Alfred J. Tuchfarber

THE POLLS: The European Peace Movement and the Deployment of Nuclear Missiles
Connie de Boer

NEWS AND NOTES · BOOK REVIEWS
BOOK NOTES

SPRING 1985

Early Calls of Election Results and Exit Polls: Pros, Cons, and Constitutional Considerations

The use of exit polls has made possible early, accurate projections of winning candidates on Election Day, especially in landslide elections. But the saturation of the mass media with exit poll results can produce a backlash of adverse public reactions, ranging from hostility toward polling to political apathy and nonvoting. Several states have sought to discourage exit polling by laws prohibiting ready access to voters as they leave polling places. One such law in the state of Washington was challenged in the federal court in Tacoma. It was adjudged constitutional in July 1984, but the decision was reversed on appeal just prior to the 1984 election. The issue remains unresolved as POQ goes to press. The controversy over exit polling, which has implications for both the news media and the opinion research profession, was discussed in a plenary session of the 39th Annual AAPOR/WAPOR Conference. The text that follows is an edited, condensed version of the transcripts of that session, held on May 18, 1984 in Delavan, Wisconsin.

Introduction

J. RONALD MILAVSKY

How soon should the winners and losers of political races be reported? Due in large part to the development of the technology of survey research as applied to exit polls, it is now possible to know with a high degree of certainty who are the winners and losers quite early on election day, well before polls close in a large number of states. From a research point of view this is indeed a wonderful accomplishment.

The greater the margin between the winner and the loser, the sooner in the day it is possible to announce. Conversely, the smaller the margin between the winner and loser, the later in the day it is possible. When calls are made late, as they were in the presidential election of 1976, there is no problem. Problems may arise when calls are made before the polls close in many states. There is at least a possibility that the behavior of some voters, particularly those in western states, may be affected.

J. Ronald Milavsky, Vice-President for News and Social Research at NBC, served as AAPOR Chairman in 1983–84. He organized and moderated this plenary session.

Public Opinion Quarterly Vol. 49:1–18 © by the Trustees of Columbia University
Published by Elsevier Science Publishing Co., Inc. 0033-362X/85/0049-1/$2.50

As professional survey researchers, we have a stake in this issue. A law in the state of Washington, which may become a model for similar laws in other states, aims at making valid exit polls impossible to conduct. As professionals, we have to wonder whether legal restrictions in this specific, political context will spawn further legal restrictions on survey research in other contexts.

This discussion focuses on policy issues, rather than methodological ones, in particular on these questions: Since the technology is available, should it be used to report what is known when it is known, or should that information be held until some later point in time? Indeed, in this age of instant communication via satellites, can it be held, and for how long? Each of the four members of this panel has spent a great deal of time thinking about this issue, and each has something different to say about it.

The Congressional Concern About Early Calls

AL SWIFT

Here is an issue on which there are many opportunities for honest disagreement. One of the problems is that there is very little solid research on what the effect of early projections is. When we argue about the effects that early projections have we are arguing thoughtful opinion rather than solid research. On my side of the argument are the true believers, on the other side of the argument are the atheists, and neither of us can prove our point.

In Congress we have been trying desperately to get some voluntary compliance to withhold projecting returns until the polls are closed. "Voluntary" is the key word. What we wanted to do was to raise the issue to a high enough level that we could get some attention, but not to so high that the media would feel that we were trying to harass or intimidate them. And following that process we've passed a resolution through the House Administration Committee. We informed people who needed to know about it.

But nothing happened, nevertheless. There are two possible reasons nothing happened. One is that the networks just heard and disagreed. The other is that they didn't hear at all. My sense is that there was a combination of both: they heard and didn't agree, and they heard and didn't think we were very serious about it.

However, this year I think we will pass a resolution through the full House, again calling for voluntary restraint. It will be a resolution that I

Congressman Al Swift, from the state of Washington, is chair of the House Task Force on Elections—a special subcommittee that handles all federal election laws, including laws on "premature" announcement of election returns—and is also on the Telecommunications, Consumer, and Finance subcommittees. He is a former director of news and public affairs for KVOS-TV in Bellingham, Washington.

Uncovering Rape: The Watchdog Press and the Limits of Agenda Setting

DAVID L. PROTESS, DONNA R. LEFF, STEPHEN C. BROOKS, AND MARGARET T. GORDON

THE ABILITY of the mass media to influence the general public's political, economic, and social priorities—the "agenda-setting" hypothesis—has been the subject of intensive study over the past decade. The growing body of research has not, however, yielded consistent findings. The strength of the causal relationship between media content and public agendas has been found to vary considerably from study to study, as have the methodologies used by different researchers, although both newspapers and television have produced agenda-setting effects. (See literature review by Roberts and Bachen, 1981.)

Even more inconclusive, and less researched, is the nature and extent to which the mass media influence the policy or "formal" agendas of decision makers (Downs, 1972; Molotch and Lester, 1974; Blanchard, 1974). Finally, studies that have attempted to identify linkages between media, public attitudes, and policy making processes have been at best theoretical (Molotch, et al., 1982), normative (Paletz and Entmann,

Abstract This article reports the findings of the second in a series of field experiments on the agenda-setting effects of news media investigative reports. The authors used a pretest-posttest quasi-experimental design to assess the impact of a newspaper investigative series about rape on a randomly selected group of Chicagoans and a purposive sample of policy makers. Unlike the first study, the series had a minimal impact on public opinion and policy making, but affected profoundly the subsequent newspaper coverage of rape.

David L. Protess and Donna R. Leff are Associate Professors at the Medill School of Journalism and the Center for Urban Affairs and Policy Research (CUAPR); Margaret T. Gordon is Professor at the Medill School of Journalism and Department of Sociology, and Director of the CUAPR; all are at Northwestern University, Evanston, Illinois 60201. Stephen C. Brooks is Assistant Professor in the Department of Political Science, University of Akron. The authors wish to thank Fay Cook, Peter Miller, and Tom Tyler for comments on earlier versions. Requests for reprints may be addressed to David Protess, Center for Urban Affairs and Policy Research, Northwestern University, 2040 Sheridan Rd., Evanston IL 60201.

Public Opinion Quarterly Vol. 49:19–37 © by the Trustees of Columbia University
Published by Elsevier Science Publishing Co., Inc. 0033-362X/85/0049-19/$2.50

1981; Regier, 1930), or impressionistic (Bernstein and Woodward, 1974; Crouse, 1973; Salisbury, 1980.)

This article reports the second in a continuing series of studies[1] designed to assess empirically the effects of particular news media investigative reports on public *and* formal agendas. The first (Cook, et al., 1983), utilizing a pretest-posttest experimental design, found that a nationally televised investigative news report on fraud and abuse in the federally funded home health care program had significant effects on both kinds of agendas. The study found that home health care-related issues (and not unrelated issues) became significantly more important to citizens and policy makers exposed to the television investigative report than to nonviewers. Nonetheless, actual policy changes following the televised report resulted more from direct pressure for "reform" by the journalists themselves than from demands by the general public.

The current study uses a quasi-experimental research design to measure the effects of a Chicago *Sun-Times* newspaper investigative series. The series disclosed government improprieties in the handling of cases of rape and other sexual assaults of Chicago area women. Thus, we continue to focus on media accounts of official "wrongdoing" in a social problem area. We again employ a design that is better suited to test causal hypotheses than other methodologies traditionally used in agenda-setting research (Cook and Campbell, 1979), including cross-sectional (McCombs and Shaw, 1972; McLeod, et al., 1974; Erbring, et al., 1980) or panel study designs (Tipton, et al., 1975; Shaw and McCombs, 1977; MacKuen, 1981). This study, however, uses as its content the topic of rape rather than home health care, is in a local rather than a national publication, and is in a print rather than broadcast medium.

If media indeed have an ability to influence citizen judgments of issue importance, then they might also produce other changes. For numerous reasons, most agenda-setting studies fail to pursue these possible additional consequences. This article first will examine the attitudinal impact of the *Sun-Times* rape series. Next we trace the effects of the series on public policy making in Chicago. Finally, we will measure the effects of the series on the *news media's own agenda of concerns*: specifically, subsequent *Sun-Times* coverage of rape and related issues.

Research Design

The pretest-posttest experimental design is highly appropriate, but not traditionally utilized, in research involving nonlaboratory studies of media agenda-setting. One significant reason for this is the nature of news. The inability to determine in most cases when a news media story is about to break makes pretesting and experimental or quasi-experimental manipulation difficult, if not impossible, to achieve.

---

[1] The studies are being conducted by an interdisciplinary research team at the Center for Urban Affairs and Policy Research, Northwestern University.

The use of a quasi-experimental design in this study is attributable to two unique characteristics of this research. First, the newspaper series we studied is investigative in nature; reporters worked for several months to unearth evidence of wrongdoing by public officials. While the publication of an investigative series may appear to newspaper readers (or television viewers) to be breaking news, in fact the often lengthy preparation time makes advance knowledge of such stories possible and thus pretest surveys practicable. Further, the policy relevance of most investigative reports makes them particularly suitable for studying their effects on both citizen and formal policy making agendas.

The second distinctive feature of our study is the cooperation of reporters who informed researchers of the content of the investigative series months before its publication. Thus, surveys of respondents could be made both before and after the series was published, about issues specifically related and unrelated to its subject matter.

The investigative series, "Rape: Every Woman's Nightmare," was published *exclusively* by the Chicago *Sun-Times* in five parts beginning on July 25, 1982. Each part was headlined on the front page of the tabloid newspaper,[2] with the text beginning in the first five pages. The newspaper reported that rape and other assaults of Chicago area women were far more commonplace than police statistics indicated; that some local police departments intentionally were underreportng the crime; that several Chicago suburbs had higher incidences of rape than some inner city neighborhoods; and that state and local officials were not taking adequate measures to punish offenders and assist victims.

Respondents interviewed by our research team about these allegations included a random sample of the general public and a purposive sample of policy makers. The methods used in drawing the samples and analyzing the data are described below.

GENERAL PUBLIC

Three-hundred-forty-seven respondents from the Chicago Metropolitan area were selected from a pool of potential interviewees using random digit-dialing techniques. Telephone numbers were generated from listings taken from area directories. The last digit of each chosen number was replaced with a random digit to insure that unlisted telephone numbers were included in the sample. Interviewing took place June 1–4, 1982. The response rate was 71 percent; those who refused to be interviewed were found to be similar to respondents in sex, age, education, and area of residence (city or suburbs).

---

[2] The headlines of the five-part series were: "Rape Epidemic: 'No Woman Immune' . . . Sex Crime Rates by Neighborhood;" "Many Never Face Trial in Rape Cases;" "Help for Rape Victims Tough to Find . . . Funds Cut Off for Many Service Programs;" "Prisons Fail to Aid Rapists."

Since the *Sun-Times* series focused on the problems affecting primarily women and suburbanites, the sample was stratified to reflect proportionately their presence in the general population. The sample was further stratified by readership. Respondents were asked at the outset of the interview if they "regularly read or subscribe" to the *Sun-Times*, the Chicago *Tribune* (the *Sun-Times'* major competitor), or other newspapers. Of the sample of randomly selected respondents, 170 identified themselves as *Sun-Times* readers, while 177 were readers of other newspapers and *not* the *Sun-Times*.

The entire sample was recontacted one week after publication of the newspaper series. As shown in Table 1, 187 of the respondents agreed to be reinterviewed; interviews were conducted August 2–6, 1982. These respondents had similar demographic characteristics to those not reinterviewed. It is this final pool of 187 respondents who comprised the quasi-experimental and control groups in the study.

Two criteria were used to determine whether respondents actually had been exposed to the newspaper series. First, expecting that regular *Sun-Times* readers would be exposed to the series, we defined them as a treatment group for the purpose of data analysis. The non-*Sun-Times* readers were defined as a "control" group. Table 1 gives the distribution of respondents between the two groups.

Since "cross-pollination" of important news stories between media outlets is a common phenomenon, we further anticipated that some

Table 1. Quasi-Experimental Design Distribution of Respondents

	Pretest		*Posttest*	
Respondent Group	*N*	%	*N*	%
Readers/nonreaders				
Readers (treatment)	170	49.0	85	45.5
Nonreaders (control)	177	51.0	87	46.5
Unclassified cases	00	0.0	15	8.0
Total	347	(100.0%)	187	(100.0%)
Series aware/not aware				
Series aware (treatment)	NA		125	66.9
Not aware (control)	NA		55	29.4
Unclassified cases	NA		7	3.7
Total	NA		187	100.0%

respondents may have become aware of the series even if they had not read the *Sun-Times*. To determine whether this had occurred, respondents were asked *at the end of the posttest interview* if they had "seen, read, or heard anything about recent news stories concerning serious crime, particularly rape, in the metropolitan Chicago area?" For analytic purposes, we labeled those who responded yes to this question as a second treatment group, and no as a control group. The distribution of respondents for this analysis is also shown in Table 1.[3] It is important to emphasize that respondents in both analyses are from the same overall sample. It is only the quasi-experimental definition of "treatment" and "control" that changes.

In the pre- and posttest interviews, respondents were asked questions related to crime in general and to rape and assaults of women in particular (see Appendix). (These were embedded among questions on other social issues; see below.) On the general level, respondents were asked to list "the most critical urban problems facing us today." Specifically, they were asked to assess the importance of lenient treatment of accused rapists, and of providing financial assistance to rape crisis centers. The agenda-setting hypothesis would, of course, expect statistically significant changes in the salience of these issues in the treatment groups.

Respondents also were asked to rate the importance of other issues irrelevant to the *Sun-Times* series, including inadequate day care services, fraud and waste in government programs, poor government regulation of medical institutions, and the desirability of providing financial assistance to other kinds of crime victims. Of 40 issue-related items asked in the survey, 7 (18 percent) concerned noncrime-related issues, 28 (70 percent) concerned general crime issues or crimes other than rape, and only 5 (13 percent) specifically addressed the problem of rape.

The broad range of questions was designed to introduce further control into the research design. We hypothesized pre- to posttest consistency on the "irrelevant" items, in comparison with the agenda-setting changes expected for items related to the content of the newspaper series. A further purpose of asking unrelated questions was to reduce respondent sensitivity to the rape issue. While the pretest may have alerted respondents to be more aware of certain issues, the broad variety of questions reduced any focused sensitivity.

[3] An additional analysis of those who were aware of the series was also conducted. Further questions were asked of those who responded yes to the awareness question to determine the source of this awareness. The breakdown of the 187 "series aware" readers was as follows:

From the *Sun-Times*	68
From other than the *Sun-Times*	57
Unclassified cases	62
Total	187

In the crime-related questions, respondents were asked to compare the prevalence of rape and other crimes in the city and suburbs, to assess law enforcement effectiveness in dealing with rape and other crimes, and to state their views on police discretion in categorizing and reporting rape and other crimes. Finally, respondents were asked personal background questions about their fear of crime and personal victimization experiences. As a possible explanatory factor for evaluating the results of the survey, respondents also were questioned about their views of the reliability of news media crime reporting.

POLICY MAKERS

A purposive sample of 39 policy makers was interviewed by telephone before the newspaper series was published. Thirty-seven (94.9 percent) were reinterviewed one to two weeks after publication. The analyses to be described are based on these pre- and posttested respondents.

Selection of the policy makers was not based on their newspaper readership habits, since it was difficult to find nonreaders of a front-page crime series in this particular population group. In fact, 36 of the 37 policy makers indicated in posttest questioning that they had "seen, read, or heard (something) about recent news stories concerning serious crime, particularly rape, in the Chicago Metropolitan area."

The main criteria for selecting the policy maker sample were that they hold responsible positions in local government or private agencies that significantly affect the criminal justice system, and that they had no advance knowledge of the forthcoming newspaper series. The policy makers primarily were local government officials, including mayors, city managers, suburban village presidents, city and suburban police chiefs, and criminal prosecutors in the State Attorney's office. Several representatives of victim/witness programs and women's advocacy groups also were interviewed.

The policy makers were asked a series of questions that were identical to those asked of the general public. Questions about the importance of various problems related to rape were included, as were items not related to the content of the newspaper series. Further, the policy makers were asked to estimate the *public's* views of the importance of these issues, and to give their own views of the accuracy of the news media in general, and investigative reporting in particular.

The before-after aspect of the questioning allowed the reseachers to determine possible agenda-setting effects of the series on policy makers, while the nature of the questions allowed us to compare the impact with that of the general public. Posttest questions about what they intended to do about the series' findings and additional follow-up interviews with an expanded sample of policy makers facilitated an analysis of media impact on actual policy making.

NEWS MEDIA

A generally unexplored area of mass media research is the impact that news media reports have on media themselves. We suggest that it is appropriate to attempt to measure agenda-setting on the *media*, and in particular to hypothesize about the impact of investigative series on future news judgments about the subject of such reports.

The *Sun-Times* was analyzed each day for a three-month period prior to the publication of the rape series, and for three months after publication, excluding stories that obviously were follow-up stories to the original series. For comparative purposes, a content analysis also was performed of the Chicago *Tribune* during the same time period.

The content analyses covered news reports of incidents of rape or sexual assault, and official actions related to rape, e.g., arrests, prosecutions, and legislative initiatives. Researchers counted the total number of articles, column inches, and page placement of the articles before and after publication of the investigative reports. In addition, an assessment was made of the actual content of the articles, in part directed at finding whether the newspapers utilized the investigative revelations to provide more analytical and contextual coverage of rape following publication of the series.

Findings

GENERAL PUBLIC

Our major concern here was whether the *Sun-Times* investigative series changed public opinion about rape and related issues. Specifically, we attempted to measure agenda-setting changes among those exposed to the series, and to make comparisons with the unexposed group of respondents. We prepared two analyses using different definitions of exposure to the series, as discussed above.

Table 2 provides the mean response to survey questions asked before and after the publication of the series.[4] Responses are categorized according to the two types of treatment and control criteria. The table includes responses to six sets of questions about issues related to rape and other social problems. The first three questions were designed to measure attitude changes about the importance of the problem of rape, and possible corrective actions that might be taken in the policy arena. The second set of questions were designed to measure what readers had "learned" about rape from points made in the newspaper series.

---

[4] In addition to the mean responses, we examined also the total *number* of responses moving toward and away from the "agenda-setting" hypothesis. Of the 425 responses by "*Sun-Times* readers" to the five specifically rape-related items, 98 (23.1 percent) changed in the direction hypothesized (toward a score of 1), while 69 (13.9 percent) moved toward a score of 4. The others (63.0 percent) did not change. Responses of *Sun-Times* readers to the items unrelated to rape were random and largely offsetting, as were the responses of the non-*Sun-Times* readers to both related an unrelated items. Of the 625 responses by "Series

When we examine these results for Readers/Nonreaders, it is evident from the measures we used that the investigative series had highly limited agenda-setting effects on *Sun-Times* readers. The only statistically significant change on a rape-related item occurred in response to the open-ended question about "critical problems facing Chicago." "Crime" was more likely to be listed by *Sun-Times* readers in the posttest as one of the two most critical problems facing the city, while similar change did not occur for non-*Sun-Times* readers. However, this agenda-setting effect is highly general in nature, and may not be related to the specific content of the investigative report about rape. None of the questions related to the specific content of the newspaper series itself yielded a significant change of opinion or knowledge about rape or related issues.

The same pattern prevailed in the Series Aware/Not Aware groups. Series Aware respondents were significantly more concerned in the posttest about the importance of crime in general as a problem facing Chicago, whereas respondents who were unaware of the series did not experience as strong a change. The treatment group failed to change its views on any matters specifically related to the content of the rape series.[5]

In contrast, the public changed significantly on some crime-related issues *not* raised by the rape series. Responses related to drug abuse and drug crimes had the highest number of statistically significant changes. Other issues did not produce consistently significant responses. Since only 10 of the 63 tests in Table 2 produced changes that were at the .05

Aware" respondents to the five specifically rape-related items, 135 (21.6 percent) changed in the direction hypothesized, while 108 (15.7 percent) moved toward a score of 4. The others (62.7 percent) did not change. Series Aware responses to the items unrelated to rape were random and largely offsetting, as were Series Not Aware responses to both related and unrelated items. Looking at the number of changes on individual items does not alter the conclusion that changes among both readers and nonreaders, or among the aware and the nonaware, were essentially random.

[5] Further analysis of subcategories of the Series Aware group produced a similar lack of change (see footnote 3 for distribution of this group). However, respondents who were exposed to the series' findings by reading the *Sun-Times* did become significantly more concerned that the government provide financial support for rape crisis centers. The newspaper series had described a major controversy surrounding the provision of such funds, and editorially supported the position that they be increased. Since this is the only "learning" change that took place, its importance should be viewed with caution.

Table 2. Mean Responses for General Public[a]

	Readers/Nonreaders				*Series Aware/Not Aware*			
	Treatment (Readers) (N=85)		*Control (Nonreaders) (N=87)*		*Treatment (Aware) (N=125)*		*Control (Not Aware) (N=55)*	
	Pre	*Post*	*Pre*	*Post*	*Pre*	*Post*	*Pre*	*Post*
Two most critical problems (1=crime as one, 2=crime not mentioned	1.59	1.40*	1.59	1.48	1.57	1.40*	1.66	1.55
Importance of problem								
Lenient treatment of rape	1.21	1.19	1.22	1.18	1.24	1.22	1.14	1.11
Inadequate day care	1.88	1.75	1.98	1.98	1.97	1.86	1.84	1.86
Fraud and waste in government	1.41	1.34	1.35	1.30	1.42	1.36	1.28	1.21
Poor health regulations	1.61	1.54	1.74	1.72	1.69	1.70	1.62	1.50
Importance of government help								
Rape crisis centers	1.31	1.22	1.41	1.36	1.38	1.29	1.34	1.32
Battered wife shelters	1.55	1.46	1.41	1.36	1.66	1.54	1.58	1.37
Abused children protection	1.17	1.13	1.27	1.16*	1.30	1.16*	1.06	1.09
Theft and burglary victims	2.34	2.28	2.50	2.41	2.44	2.33	2.43	2.32
Location of incidents								
Rape more in suburbs	3.11	3.17	3.29	3.15	3.22	3.24	3.14	3.04
Property crime more in suburbs	2.85	2.85	2.94	2.87	2.93	2.89	2.83	2.83
Child abuse more in Chicago	2.26	2.20	2.40	2.48	2.31	2.28	2.37	2.57
Drug crime more in Chicago	2.53	2.21*	2.59	2.39	2.51	2.32	2.72	2.30*
Effective law enforcement programs								
Rape and sexual assault	2.94	2.97	3.20	3.08	3.05	3.03	3.06	3.02
Drug abuse	2.95	2.86	2.90	2.80	2.89	2.90	2.98	2.74
Juvenile gang control	3.12	3.17	3.20	3.08	3.05	3.05	2.96	2.94
Drunk driving	2.72	2.89	3.04	2.80*	2.89	2.90	2.89	2.72
Crime the same as police statistics								
Rape	1.31	1.25	1.40	1.31	1.38	1.30	1.32	1.28
Homicide	1.56	1.57	1.52	1.57	1.63	1.60	1.33	1.49
Aggravated assault	1.41	1.37	1.61	1.48	1.65	1.43*	1.27	1.45
Drug related death	1.48	1.50	1.50	1.48	1.60	1.41*	1.25	1.73*

[a] Lower numbers indicate more positive responses (i.e., 1=very important; 4=very unimportant).

Significance of change from pretest to posttest using a dependent t=test, $p<.05$.

level, no definitive conclusions can be drawn. Finally, responses to questions asked about issues entirely unrelated to problems of urban crime were not included in the tables because, as expected, no statistically significant changes occurred.

Overall, the data suggest that the *Sun-Times* series of investigative reports had a minimal impact on public attitudes concerning rape, but may have had somewhat more effect on nonrape crime issues and general concern about crime. This is in marked contrast to the agenda-setting effects identified in our first study. Possible explanations for this difference are discussed below.

POLICY MAKERS AND POLICY MAKING

As described above, 37 policy makers were interviewed before and after the *Sun-Times* series to measure changes in their perception of rape-related issues, and to determine whether their assessment of public opinion had altered. Because of their continuing awareness and concern for the problem, we hypothesized that their personal views would not be affected by the newspaper series. Further, given their sensitivity to any special media attention to rape, we hypothesized that the elites would expect that public opinion toward the problem would change as a result of the front-page series.

Table 3 provides the mean responses to the questions asked of policy elites.[6] The first set of columns presents responses to questions about their own views.The second set of columns contains their opinions of the public's attitudes about these questions. Only those questions relating to information provided in the series were asked in this second way.

The data again show that the publication of the *Sun-Times* series had little impact on either elite's personal views, their work routines, or their assessment of public opinion. Surprisingly, however, the greater mean changes and the only statistically significant change on rape-related questions were for personal attitudes. In particular, the elites believed at the posttest interview phase that law enforcement programs regarding rape and sexual assault were significantly *less* effective. It seems that elites were influenced by the series on this one issue, but that they did not believe that it would have the same impact on the general public.

On the basis of our analyses of legislative hearings, bills, statutes, budgets, news releases, and interviews with the political actors who "responded" to the rape series we found that policy reactions to the investigative report were largely "symbolic." There were no new budgetary or administrative changes, and only limited relevant statutory proposals accompanied rhetorical promises of "reform." The policy developments which were announced in the wake of the series had been initiated long before the *Sun-Times* investigation began; nonetheless, publication of the series gave policy makers a new opportunity to focus public attention on the problem of rape.

Thus, within a week after the first story appeared, Mayor Jane Byrne

---

[6] In addition to the mean responses, we examined also the total *number* of responses moving toward and away from the "agenda-setting" hypothesis. Of the 185 policy maker responses to the five specifically rape-related items, 33 (18.3 percent) responses changed in the direction hypothesized, while the same number (18.3 percent) moved toward a score of 4. The others (63.4 percent) did not change. Policy maker responses to the unrelated items were similarly random and offsetting.

Table 3. Mean Response for Policy Makers (N=37)[a]

	Personal View		View of Public	
	Pre	*Post*	*Pre*	*Post*
Importance of problem				
Lenient treatment of rape	1.27	1.38	Not Asked	
Inadequate day care	1.92	2.24*		
Fraud and waste in government	1.57	1.38		
Poor health regulation	1.73	1.92		
Importance of government help				
Rape crisis centers	1.54	1.49	Not Asked	
Battered wife shelters	1.76	1.65		
Abused children protection	1.30	1.22		
Theft and burglary victims	2.62	2.24		
Location of incidents				
Rape more in suburbs	3.35	3.30	3.61	3.42
Property crime more in suburbs	3.00	2.91	3.35	3.08
Child abuse more in Chicago	2.67	2.31	1.58	1.50
Drug crimes more in Chicago	2.60	2.51	1.65	1.68
Effective law enforcement programs				
Rape and sexual assault	2.32	2.70*	2.94	3.05
Drug abuse	2.54	2.51	3.32	3.11
Juvenile gang control	2.70	2.73	3.27	3.38
Drunk driving	2.76	2.62	3.24	3.03
Crime the same as police statistics				
Rape	1.68	1.49	Not Asked	
Homicide	2.03	1.92		
Aggravated assault	1.57	1.70		
Drug related death	1.95	1.73		

[a] Lower numbers indicate more positive responses (i.e., 1=very important, 1=highly effective, etc.).

* Significance of change from pretest to posttest using a dependent *t*-test, $p < .05$.

announced the "creation" of a Rape Hotline and a public housing drop-in center for rape victims, although both measures had been operational for months. Similarly, Governor James Thompson, the week the series was published, signed a bill into law providing for increased criminal penalties for "gang rape." The bill, described by its sponsors as a routine tightening of the state's criminal code, had been on the governor's desk for a month.

In the legislative arena, hearings were held and sexual assault legislation was introduced and eventually passed; both measures had been planned well before the series. Since our analysis indicates that no policy making efforts were directed at the expose's central revelation of systemic deficiencies in reporting and combating rape, it appears the illusion of reform was created without a fundamental change in opinion or policy substance. However, the series did provide another platform for those pushing for widespread reform of rape legislation.

THE PRESS

The content analyses of the *Sun-Times* showed a significant change in the depth and extent of its rape coverage after publication of the investigative series despite no change in police reports of rape. Specifically, the content analysis found:

1. An increase in the total number of stories about rape, from 32 during the three months preceding the series to 58 in the three months after the series.
2. A doubling of space devoted to rape from 594 square inches before to 1,354 after the series.
3. An increase in the number of rape stories in the first 10 pages of the newspaper from 10 to 18.
4. An increase in the number of rape stories containing interpretation, analysis, or context (rather than reporting only the facts of an individual case) from 6 to 20.
5. An increase in coverage of official governmental activities related to rape from 19 stories to 35.

In sum, it appears the most distinctive agenda-setting impact of the *Sun-Times* investigative series was on the newspaper itself. In contrast, the content analysis of the Chicago *Tribune* found a minimal change in that newspaper's coverage of rape and related issues during the same period. The total number of stories about rape increased from 39 during the three months preceding its competitors' investigative series to 41 in the three months after the series. The *Tribune* increased the space it devoted to these stories from 462 square inches (preseries) to 588 square inches (postseries.)

Suggested Explanations of Findings

GENERAL PUBLIC

The sharpest difference between our two studies of investigative reporting was in their impact on public attitudes. In the televised home health care program, results of telephone interviews showed that home health care-related issues became significantly more important to citizens and policy makers exposed to the television program than to those watching another program. The rape series showed no comparable agenda-setting effect on respondents exposed to the newspaper articles. Several explanations are offered. First, the stronger agenda-setting effects in our first study may be due to the psychological impact of broadcast media. In fact, there is growing support for the notion that the impact of television is greater than that of newspapers when the subject is dramatic and the event short-term (McClure and Patterson, 1976; Graber, 1980). Second, the tone and content of the newspaper series tended to be dry and dull compared to the television expose, which relied on undercover filming and confrontational interviews. Third, a more sensitive measure of "importance" at the upper end of the scale used in the present study may have produced clearer results.

However, it is also possible that minimal change occurred on the posttest because there may already have been so much knowledge and concern about rape at the pretest stage that there was limited possibility for the series to have an agenda-setting effect. In other words, rape as a social problem may have already reached a "saturation" level in the public's mind, and was already high on the agenda. This is to be contrasted with the home health care issue, about which little was known at the pretest stage. Consequently, in the latter case the maximum opportunity existed for the news media to influence viewer estimates of issue importance.[7]

The general public's fear of crime may also help to explain the one consistent agenda-setting effect of the rape series: on respondent's views of the increased importance of crime as a "critical urban problem facing Chicago." This finding is congruent with recent research indicating rape may be a "bellwether crime" for the public (Gordon, et al., 1980; Riger and Gordon, 1981); when rape is increasing or high on the public agenda, it seems to signal increases in personal danger of many types. Furthermore, other research indicates that news media reporting about crime may affect people's concern for it as an *issue*, without necessarily affecting their personal *fear* of particular crimes (Gordon a Heath, 1981; Tyler, 1983; Skogan and Maxfield, 1981).[8]

---

[7] When asked about the importance of home health care as a problem in pretest interviews, the mean response for experimental subjects was 1.72, where 2.00 was "somewhat important" and 1.00 was "very important." In the posttest, the mean response to this question moved to 1.54 (p .05). In contrast, the pretest response in the current study to the question about the importance of "lenient treatment of people accused of rape . . ." was 1.21, with a posttest mean of 1.19 (p .05). The intensity of feeling regarding rape is also demonstrated by the *number* of respondents expressing strong opinions at the pretest stage. For example, 151 (80.7 percent) said that law enforcement programs dealing with rape were "highly ineffective."

[8] Interestingly, the statistically significant changes on response concerning drugs and crime followed widespread media attention to these issues between the pre- and posttests. This serendipitous "effect," common in longitudinal field experiments, is nonetheless consistent with the agenda-setting hypothesis.

Finally, a variation in research design may partially account for the different findings. The first study was an experiment in which great care was taken to maximize internal validity. This was done by random assignment of respondents to treatment and control groups and the use of a selection procedure that would insure that the number of "treated" cases (those who watched the television program) was large enough for statistical analysis. Since the investigative report was aired on a commercial television show which had relatively low ratings, respondents were asked, and in some cases paid, to watch the program or an opposing program. (They were not, however, specifically directed to watch the particular segment of the one-hour show that pertained to home health care.) While this procedure provided stronger internal validity for the experiment, it threatened external validity—the ability of the results to be generalized to similar situations. The request to watch and payment may have prompted respondents to attend more closely to the message, and to give it more thought (Goetz, et al., 1984). If this is the case, it suggests that the news media's agenda-setting impact is strongest for *those who pay closest attention to their message*, as opposed to those who casually watch or skim a media report.

In contrast, the present study utilized a quasi-experimental design. Respondents were not randomly assigned, but rather placed in treatment and control groups on the basis of their self-reported readership habits, and responses to questions about their awareness of the series. This was done (1) to enhance the external validity of the study, (2) because the precise publication dates of the rape series were less certain than with the home health care piece, and (3) because of the difficulties involved in getting respondents to closely follow a five-part series.

The various tradeoffs between internal and external validity in constructing research designs are well documented elsewhere (Cook and Campbell, 1979). In terms of these two studies, the second has stronger external validity because it measures conditions as they naturally exist in society. However, the internal validity of this study may have suffered as a result.

POLICY MAKERS AND POLICY MAKING

Attitude changes of policy makers followed the pattern of change for the general public in both studies. Policy makers already were very concerned about the problem of rape and therefore we could detect no further attitude changes. As for differences between the two studies, because the *Sun-Times* was a weeklong story in one of Chicago's two newspapers, we correctly assumed that nearly every policy elite (36 of 37) would be exposed to all or part of the story by week's end, which made it impossible for us to establish a "nonexposed" series of elites, as we were able to in the home health care study.

But in the policy making arena, responses of the groups in both studies were virtually identical. swift, symbolic overtures to redress the abuses exposed in the media investigations. A crucial factor in this kind of response is the "coalition" relationship between reporter and policy maker developed in the period *before* publication to build in policy responses to the problems identified. The response was more of a collaboration in the home health care study, in part because the investigative reporters were savvy about approaching policy makers in advance (to maximize the impact of the reporters' work), whereas the rape reporters were "beat" reporters more accustomed to being uninvolved observers. The end result tended to be the same, however, because policy makers are quick to take advantage of new oportunities for "their" issues, to "cash in on some publicity," as one public official put it.[9] Thus we conclude that the impending publication of a news media investigation may cue journalists and policy makers to form symbiotic coalitions for at least symbolic reform.

THE PRESS

Perhaps the most striking change that followed the newspaper series, and the one least addressed heretofore by the agenda-setting literature, was in the *Sun-Times'* subsequent coverage of rape. No comparable data exist from the study of the television program on home health care fraud and abuse or other research of this nature. Thus, the finding that the newspaper significantly increased its own attention to rape and related issues in the months following publication of its investigative series requires separate (and somewhat speculative) treatment.

Recent research on agenda-setting (Gandy, 1982; Lang and Lang, 1981; Fishman, 1980) has identified factors that influence the construction of "media agendas," i.e., the flexible hierarchy of issues that receive fairly consistent coverage by news media institutions. We agree with the underlying premise that the news media not only can influence audiences and policy making, but can be significantly affected *by* them as well.

---

[9] Interview by David L. Protess with Francine Stein, Executive Director, Illinois House Rape Study Committee (July 1983).

creating highly interactive processes of news generation (Molotch et al. 1982). Our findings here further suggest that particular issues may rise and fall on an individual news organization's agenda because of a variety of influences, both inside and outside of the newsroom.

In particular, the findings parallel Fishman's (1980) study of a "media crime wave" against the elderly in New York City. There was a marked increase in the number of stories in the media concerning crime against the elderly, while there was a *decrease* in crimes reported to police during the same period. Fishman suggests that the media "crime wave" accounts resulted from the need of journalists to organize the mass of news events which confront them every day; he hypothesizes that they often do this by identifying *themes* in the news which then sensitize them to stories of a similar type. If and when other competing news organizations pick up on and publish stories on the theme, they may produce a *wave* of news stories on the theme topic.

In the case of the *Sun-Times* rape series, the theme of sexual victimization that developed over the lengthy investigative process appears to have sensitized reporters and editors so that they continued to keep stories about rape on their agenda for several additional months after publication. On the other hand, the relatively small increase in the *Tribune*'s coverage of rape may reflect the lack of a similar "consciousness-raising" experience. It suggests also the *Tribune*'s recognition that their media rival had at least for the time "cornered the market" of special coverage of rape with their exclusive series. Thus, the result was more of a "ripple effect" than a multimedia wave.

The basis of the change at the *Sun-Times* is further borne out by statements of journalists in postseries interviews who told researchers the changes were a result of a deliberate shift in *Sun-Times* policy. They said that their editors "suggested" to them and to beat reporters who regularly cover crime that more attention should be paid to the problem of rape. (However, Ralph Otwell, *Sun-Times* editor, asserted that any difference in rape coverage was coincidental and not related to a change in editorial direction.)[10] The reporters who did the series nevertheless said they noticed the significant differences in *Sun-Times* coverage that our content analysis identified; "it's been our main source of satisfaction of doing the series,"[11] one of the reporters said.

---

[10] Interview by Donna R. Leff with Ralph Otwell, editor, Chicago *Sun-Times* (June 1983).

[11] Interview by David L. Protess with Dennis Byrne, reporter, Chicago *Sun-Times* (February 1983).

Conclusions

This article has presented the results of a second study of the effects of an investigative news report. Unlike the first analysis, these findings show that the report had little impact, as measured, on attitudes of the public or policy makers and generated no new policies to deal with the specific abuses in the series. It did, however, provide a new platform for reannouncing initiatives and reintroducing legislation. The second study also investigated a relatively new aspect of agenda-setting and found that this investigative report had an important impact on the press's own agenda.

The disparity in findings between the two analyses leaves little room to make conclusions about the complex questions associated with agenda building. It is obvious that more work is needed. All too often, social science is willing to draw broad generalizations of complex phenomena from only one or two studies. As demonstrated here, the structure of those studies and the numerous variables involved often produce inconclusive findings. We plan to continue experimentation and quasi-experimentation of this sort with several additional investigative reports to specify further the connections between media, opinion, and policy making.

Appendix

QUESTIONS FROM THE PRE- AND POSTTEST SURVEY

1. As you are well aware, many urban problems face us today In your opinion what are the two most critical?

2. Now I would like to get your opinion on certain social problem areas: Please respond by telling me how important you think each area is. Do you feel ____ is a Very Important Problem, Somewhat Important, Somewhat Unimportant or Not at All Important.

   A. Inadequate day care services for preschool children
   B. Lenient treatment of people accused of rape or assault
   C. Fraud and waste in government programs
   D. Poor government regulation of medical institutions such as hospitals or clinics

3. The government sometimes provides financial assistance for victims of crimes. We'd like to know how important you think it is to provide this aid to victims in the following areas. Do you think providing aid for ____ is Very Important, Somewhat Important, Somewhat Unimportant or Not at All Important.

   A. Property loss due to theft or burglary
   B. Battered wife shelters
   C. Protective services for abused children
   D. Rape crisis centers

4. I'm going to read you a series of statements. Please tell me if you Strongly Agree, Somewhat Agree, Somewhat Disagree or Strongly Disagree with each.

   A. Child abuse is more prevalent in Chicago than in the suburbs
   B. Organized crime is more serious in Chicago than in the suburbs
   C. Rape and assaults against women are more prevalent in the suburbs than in Chicago
   D. Crimes against property such as burglary or vandalism are more prevalent in the suburbs than in Chicago
   E. Drug use and sale is more prevalent in Chicago than in the suburbs
   F. White collar crime is more prevalent in Chicago than in the suburbs

7. Now I'd like your opinion on the effectiveness of law enforcement agencies and programs in the following social problem areas. Do you think the law enforcement agencies and programs dealing with ___ are Highly Effective, Somewhat Effective, Somewhat Ineffective, or Highly Ineffective?

   A. Drug abuse
   B. Juvenile gang control
   C. Drunk driving
   D. Rape and sexual assault
   E. Shoplifting
   F. Crimes against property, vandalism, burglary, theft

8. Police have a lot of discretion over how they categorize and report crime in the Chicago area. Do you feel that the following crimes occur more often, as often, or less often than the police statistics indicate?

   A. Homicide
   B. Rape
   C. Robbery
   D. Burglary
   E. Aggravated assault
   F. Drug-related death

References

Bernstein, C., and B. Woodward
1974 All the President's Men. New York: Warner Books.
Blanchard, R.O.
1974 Congress and the News Media. New York: Hastings House.
Cook, F. et al.
1983 "Media and agenda-setting: effects on the public, interest group leaders, policy makers, and policy." Public Opinion Quarterly 47:16–35.
Cook, T. D., and D. T. Campbell
1979 Quasi-experimentation: Design and Analysis Issues for Field Settings. Chicago: Rand McNally.
Crouse, T.
1973 The Boys on the Bus. New York: Ballantine.
Downs, A.
1972 "Up and down with ecology: the 'issue attention cycle.'" Public Interest 28:38–50.
Ebring, L., E. Goldenberg, and A. Miller
1980 "Front-page news and real-world cues: a new look at agenda-setting by the media." American Journal of Political Science 24:16–49.
Fishman, M.
1980 Manufacturing the News. Austin: University of Texas Press.
Gandy, O. H., Jr.
1982 Beyond Agenda Setting: Information Subsidies and Public Policy. Norwood, NJ: Ablex, 1982.

Goetz, E. G., T. R. Tyler, and F. L. Cook
1984 "Promised incentives in media research: a look at data quality, sample representativeness, and response rate." Journal of Marketing Research 21.
Gordon, M. T., S. Riger, R. K. LeBailly, and L. Heath
1980 "Crime, women and the quality of urban life." Signs: Journal of Women in Culture and Society 5:S144–S160.
Gordon, M. T., and L. Heath
1981 "The news business, crime and fear." In Dan A. Lewis (ed.), Reactions to Crime. Beverly Hills, CA: Sage Publications.
Graber, D.
1980 Mass Media and American Politics. Washington, D.C.: Congressional Quarterly Press.
Lang, G., and K. Lang
1981 "Watergate: an exploration of the agenda-building process." In G. C. Wilhoit and H. deBock (eds.), Mass Communication Review Yearbook. Beverly Hills: Sage.
MacKuen, M. B.
1981 "Social communication and the mass policy agenda." In M. B MacKuen and S. L. Coombs, More than News: Media Power in Public Affairs. Beverly Hills: Sage.
McClure, R., and T. Patterson
1976 "Print vs. network news." Journal of Communication 26:23–28.
McCombs, M., and D. L. Shaw
1972 "The agenda-setting functions of the mass media." Public Opinion Quarterly 36:176–87.
McLeod, J. M., L. B. Becker, and J. E. Byrnes
1974 "Another look at the agenda-setting function of the press." Communication Research 1:131–66.
Molotch, H., and M. Lester
1974 "News as purposive behavior: on the strategic use of routine events, accidents and scandals." American Sociological Review 39:01–12.
Molotch, H., D. Protess, and M. T. Gordon
1982 "The media-policy connection." Working Paper Series, Northwestern University Center for Urban Affairs and Policy Research.
Paletz, D., and R. Entman
1981. Media Power Politics. New York: Free Press.
Regier, C. C.
1930 Era of the Muckrakers. Chapel Hill: University of North Carolina Press.
Riger, S., and M. T. Gordon
1981 "Fear of rape." Journal of Social Issues 37:71–92.
Roberts, D. F., and C. M. Bachen
1981 "Mass communication effects." Annual Review of Psychology 32:307–56.
Salisbury, H. E.
1980 Without Fear or Favor. New York: Times Books.
Shaw, D. L., and M. E. McCombs (eds.)
1977 The Emergence of American Political Issues: The Agenda-Setting Function of the Press. St. Paul: West Publishing Company.
Skogan, W., and M. Maxfield
1981 Coping with Crime. Beverly Hills: Sage Publications.
Tipton, L., R.Haney and J. Basehеart
1975 "Media agenda-setting in city and state election campaigns." Journalism Quarterly 52:15–22.
Tyler, Tom R.
1983 "Assessing the risk of crime victimization: the integration of personal victimization experience and socially transmitted information about crime." Journal of Social Issues, special number on criminal victimization.

Television News, Real-World Cues, and Changes in the Public Agenda

ROY L. BEHR AND SHANTO IYENGAR

STUDENTS of public opinion have long sought to understand the dynamics of citizens' concerns for political issues. The public agenda is remarkably volatile—problems surge and decline in rapid succession as focal points of public attention and concern. Following the lead of Lippman (1922), researchers have argued that changes in the public agenda are caused by changes in the media agenda. According to this agenda-setting hypothesis, the amount of media attention devoted to particular issues determines the degree of public concern for these issues (for recent reviews of the vast literature on agenda-setting, see MacKuen and Coombs, 1982; Erbring, et al., 1980).

That media agendas determine citizens' political priorities has important political ramifications. By raising public concern for particular issues, the media may also alter the criteria citizens use to evaluate their leaders. As members of the Carter Administration have suggested (see Public Opinion Magazine, 1981), heavy television coverage of Iran im-

Abstract This paper examines the interrelationships between real-world cues, television news coverage, and public concern for the issues of energy, inflation, and unemployment. On the basis of longitudinal data, the authors show that media agenda setting is indeed unidirectional—television news influences public concern and not vice versa. Lead stories are significantly more powerful than ordinary stories in shaping the public's agenda. Prevailing conditions and events affect public opinion both directly and indirectly, by determining the degree of news coverage accorded issues.

Roy Behr is a Ph.D. candidate in political science at Yale University. Shanto Iyengar is Associate Professor of Political Science at the State University of New York at Stony Brook. The authors are grateful to Bill Adams for providing research facilities at the George Washington University Library, and to Steven Rosenstone who contributed significantly to the preparation of this paper. An earlier version of this paper was presented at the 37th Annual Meeting of the American Association of Public Opinion Research, Hunt Valley, Maryland, May 20–23, 1982.

Public Opinion Quarterly Vol. 49:38–57 © by the Trustees of Columbia University
Published by Elsevier Science Publishing Co., Inc. 0033-362X/85/0049-38/$2.50

mediately before the election may have increased the importance voters assigned to foreign policy as a criterion for evaluating the presidential candidates. The election was thereby transformed into a referendum on the Carter Administration's foreign policy, with disastrous consequences for Carter's candidacy. Experimental studies tend to corroborate this argument; individuals exposed to more news about a particular issue come to assign greater weight to this issue when evaluating the incumbent president (see Iyengar, et al., 1984), and when formulating their voting choices (Iyengar and Kinder, 1984; Behr, 1984).

In this article we rely on longitudinal data spanning seven years to investigate the public's concern for inflation, unemployment, and energy. Our analysis extends agenda-setting research into two hitherto unexplored areas. First, we trace public concern to both television news coverage and to prevailing conditions or events. Second, we investigate the impact of television news on public concern as well as the reciprocal impact of public concern on levels of news coverage.

Limitations of Past Research

Given the political ramifications of agenda-setting and the preeminence of television as a credible and trustworthy source of political information (see Bower, 1983), it is not surprising that a good deal of research has been directed at the agenda-setting effects of television news programs. These studies report a kaleidescope of findings. Some researchers claim that television news coverage has no impact on the audience's perceptions of issues (Patterson and McClure, 1976); others suggest that television lags behind newspapers as an agenda-setter (Benton and Frazier, 1976) and that the effects of television on perceptions of issue importance are limited to the least educated and informed segments of the citizenry (McCombs, 1976).

Confusion over the agenda-setting power of television may simply be a product of methodological limitations. The cross-sectional sample survey favored by most researchers is hardly a powerful means of testing a dynamic process such as agenda-setting. A more appropriate strategy is to search for media effects over time, as news coverage and public concern evolve. Longitudinal studies of agenda-setting, though few in number, find that changes in the level of media attention do indeed produce changes in public concern for national issues (see MacKuen and Coombs, 1982). Similarly, experimental work demonstrates that network newscasts possess an uncanny capacity to shape viewers' political concerns (Iyengar, et al., 1982). In short, it is premature to dismiss television as a medium with little agenda-setting clout.

Most research on agenda-setting is not only methodologically weak, it also suffers from conceptual limitations. Researchers ignore the effects of direct experience on individuals' political concerns (two exceptions are MacKuen and Coombs, 1982; Erbring, et al., 1980). While there can be no denying that citizens are highly dependent upon the media for public affairs information, personal experience too is a sufficiently credible source of information. Many national issues impinge on large numbers of individuals; some issues, including crime, civil rights, and unemployment have profound personal significance. Explanations of the "issue-attention" cycle must therefore include both mediated and direct experiences; ignoring the latter may lead to notably exaggerated estimates of the effects of media.

Prevailing circumstances and events can also affect individuals' political concerns indirectly by determining what the media pay attention to. As unemployment rises, the media may devote more time to unemployment and the public's concern for the issue rises. If this is the case, then agenda-setting represents the media alerting citizens to current realities. On the other hand, despite the claims of network executives that their newscasts are "mirror images" of current realities, news coverage of national issues may be quite indifferent to prevailing conditions. Media agendas may be determined instead by idiosyncratic editorial, organizational, or commercial imperatives (see Epstein, 1973; Altheide, 1976; Gans, 1980), thus diverting the public from the "real" problems facing the nation.

In either event, it is imperative that indicators of national conditions be brought to bear on the relationship between news coverage and issue salience. Real-world indicators serve two purposes: first, to assess the sensitivity of television news coverage to current conditions and events; second, to distinguish between the effects of news coverage and real-world conditions on public concern for issues.

Finally, agenda-setting researchers have generally ignored the critical question of causality. It is taken for granted that news coverage is the driving force and that agenda-setting is a unidirectional or recursive process. The possibility of a feedback effect, namely, that public concern itself spawns news coverage, is ignored. The networks may choose to broadcast stories that are of current interest to their viewers. Our analysis is sensitive to this possibility. We attempt to estimate the impact of news coverage on public concern as well as the impact of public concern on news coverage.

Indicators

Researchers seeking to study the salience of national problems face a number of methodological difficulties. The most serious stems from the fact that no survey organization *regularly* asks some version of the question "What do you think is the most important problem facing the nation today?" Longitudinal analyses of issue salience require frequent observations, for the public's issue concerns are notably volatile. For example, in November 1974, 9 percent of the responses expressed a concern with unemployment; two months later, the figure was 25 percent. In addition, estimation of the time span over which a given variable affects issue salience requires that observations be made at regular intervals.

Our study meets these requirements by relying on three different sources of data on problem salience—Gallup, Yankelovich, and University of Michigan surveys. Each of these surveys asks a similarly worded open-ended question on the importance of national problems. Our method for making the results of the three organizations compatible is described in Appendix A, along with a list of the surveys used. In essence we calculated the percentage of *responses*, rather than *respondents* citing each issue. Combining the three sets of polls, we obtained a measure of problem importance for every two-month period between 1974 and 1980, save three. Figures for those three periods were interpolated, resulting in a total of 42 bimonthly observations.

As our indicators of television news coverage, we counted the total number of stories as well as the number of lead stories aired on the weekday CBS national news for the same period—1974–80.[1] It is useful to distinguish between lead and other stories, because the networks carefully select the former as the most important news development of the day. Lead story coverage may be more sensitive to real-world conditions and may exert a stronger impact on viewers' perceptions of political issues.

Only stories dealing with energy, unemployment, or inflation were compiled. News items that were less than 20 seconds in length were ignored. The number of ordinary and lead stories was tallied for each month and averaged for each bimonthly observation.[2]

Within each issue area, we collected indicators of current conditions. These indicators reflect the economic health of the nation. For inflation they included the overall consumer price index (cpi), the cpi for food, and interest rates. The indicators of unemployment were the unemployment rate and the average duration of unemployment. Finally, in the case of energy, we included the energy cpi, energy imports, energy stocks, home heating oil prices, and gasoline prices.

---

[1] We chose CBS for it has the largest audience. In fact, the network newscasts are so homogeneous that bias stemming from the choice of any particular network is minimal (see Meadow, 1973).

[2] Our source was the Vanderbilt Television Archive's *Abstracts* of the daily evening newscasts. News stories were coded on the basis of their principal subject matter. That is, a story that referred to two of the issues under investigation was classified according to its predominant issue content.

To supplement these indicators of current conditions, we also recorded instances of presidential addresses to the nation on each issue. It is common for presidents to signal their concern for an issue by speaking to the nation on the subject. Does presidential attention to an issue induce the media and the public to accord the issue greater importance?[3]

The networks are also likely to cover the activities of other visible political actors. In the case of energy, the OPEC oil ministers represent a group whose decisions have vital consequences for energy prices and supply in the U.S. We therefore recorded the dates of OPEC meetings.[4]

Analysis

With our indicators of issue salience, real-world conditions, and television news coverage in hand, we can attempt to specify a two-stage model of agenda-setting. The model estimates the effects of real-world conditions on news coverage and isolates the independent effects of television news and real-world conditions on issue salience. Finally, the model estimates the feedback effect of issue salience on news coverage. (A more detailed description of the model is provided in Appendix B.)

Prior to estimating the two-stage model, it is helpful to display the simple relationships between problem salience, news coverage, and real-world conditions. Figures 1–3 plot salience, total news coverage (lead and nonlead stories), and an indicator of current conditions for each issue. The indicators are percentage change in the energy cpi, the unemployment rate, and, for inflation, the percentage change in the cpi. The ranges of each variable are shown in the figures.

Unemployment is the least covered of the three issues. CBS broadcasts an average of 3.6 unemployment stories per month, as compared with 12.6 for energy and 16.6 for inflation. Inflation is the most salient issue

---

[3] The subjects of presidential addresses were obtained from the journal *Vital Speeches*. Only addresses to the nation were coded. Dates of speeches on energy were: April 18, 1977, November 8, 1977; January 19, 1978; April 15, 1979; July 15, 1979; and March 14, 1980. Dates of speeches on unemployment were: January 15, 1974; January 19, 1976; January 12, 1977; February 2, 1977; and January 19, 1978. Dates of speeches on inflation were: July 25, 1974; August 12, 1974; October 24, 1978, January 23, 1979; and March 14, 1980.

[4] Only meetings of OPEC *ministers* were coded. The dates of these meetings were: January 19, 1974; September 13, 1974; December 13, 1974; March 5, 1975; June 10, 1975; September 28, 1975; December 20, 1975; May 26, 1976; December 15, 1976; December 19, 1977; June 19, 1978; December 15, 1978; March 26, 1979; June 27, 1979; December 17, 1979; May 6, 1980; June 9, 1980; September 16, 1980; and December 15, 1980.

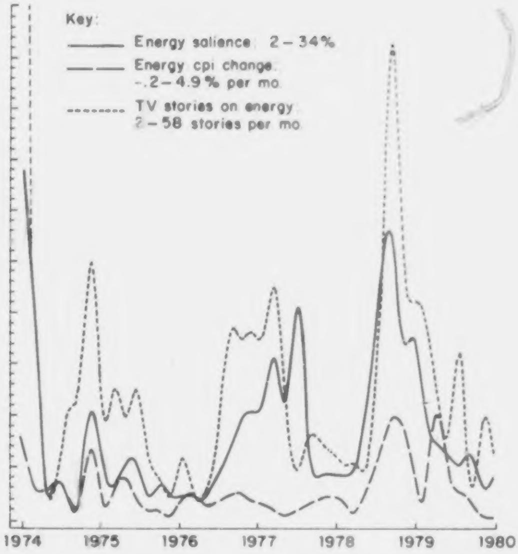


Figure 1. Energy: Percent of public naming energy as most important problem, number of TV stories on energy, percent change in energy CPI, 1974–1980

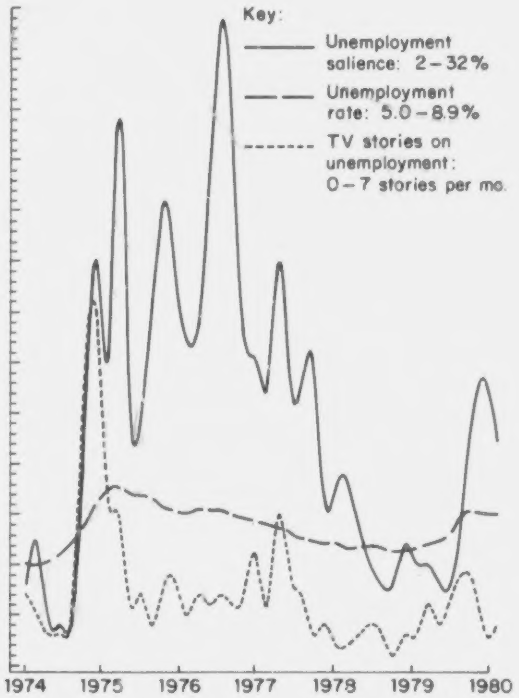


Figure 2. Unemployment: Percent of public naming unemployment as most important problem, number of TV stories on unemployment, unemployment rate, 1974–1980

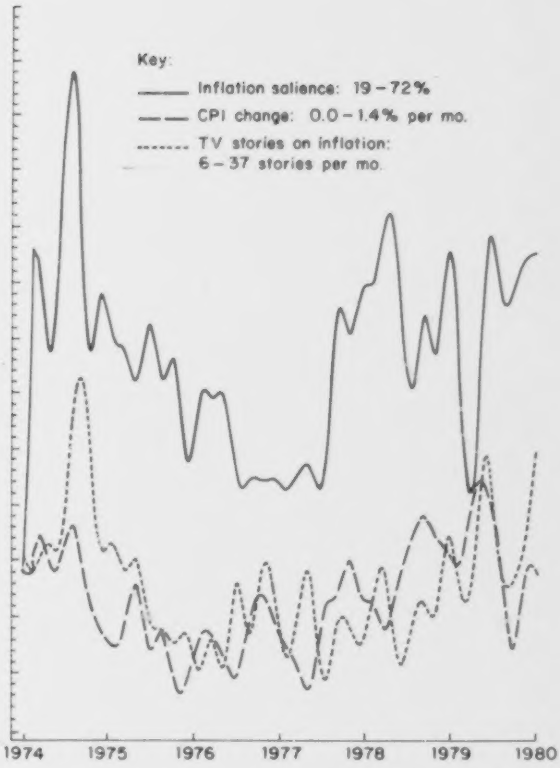


Figure 3. Inflation: Percent of public naming inflation as most important problem, number of TV stories on inflation, percent change in CPI, 1974–1980

over the seven-year span and receives an average of 41.6 percent of the responses to the "most important national problem" question. Energy and unemployment are considerably less salient, with means of 8.5 percent and 12.2 percent, respectively.

Within all three issue areas there is substantial congruence between news coverage and public concern. The simple correlation *(r)* between the number of news stories and the level of concern is .73 for energy, .37 for unemployment, and .53 for inflation. The correlations between the level of public concern and the economic indicators are also substantial—.52 in the case of energy, .74 for unemployment, and .43 for inflation.

Having seen that public concern, news coverage, and real-world conditions are intertwined within each area to differing degrees, we now proceed to identify the causal links between these factors. For purposes of organizational simplicity we present the results in two sections—determinants of news coverage and determinants of the public agenda.

Determinants of News Coverage

In this section we assess the impact of economic conditions, presidential speeches, OPEC meetings (in the case of energy), and the level of public concern on network news coverage of inflation, unemployment, and energy. We shall examine total and lead story coverage separately so as to investigate whether the two types of coverage respond to the same forces. The estimated coefficients for the total number of stories and the number of lead stories are presented in Tables 1 and 2, respectively.

In all-three areas, both total coverage and lead story coverage are affected significantly by national economic conditions. Energy coverage responds to changes in the energy cpi, coverage of unemployment is spurred by both the rate of unemployment and changes in the rate, while stories about inflation follow unusually large increases in the cpi.

These effects are substantial. A large increase in the cost of energy, for instance (6 percent in two months), produces an average of 21 stories in that two-month period. Approximately four of these will be lead stories. A shift of .25 percent *in either direction* in the level of unemployment

Table 1. Determinants of Total News Coverage

Variable	*Energy*[a]	*Unemployment*[b]	*Inflation*[c]
% Change in energy cpi	3.50 (.87)		
Unemployment rate		1.83 (.70)	
Change in unemployment rate		17.02 (2.79)	
% Change in cpi (above mean only)			3.99 (2.28)
Presidential speech	9.99 (4.39)		
OPEC meeting	7.38 (2.74)		
Public concern for issue			.67 (.36)
Constant	8.85 (4.43)	−9.96 (4.77)	1.21 (14.94)
Rho	.52 (.14)	 (.14)	.47
Adjusted R^2	.37	.58	.24
S. E. of regression	10.99	4.27	10.76
Durbin-Watson statistic	1.85	1.90	

[a] Maximum likelihood estimates.
[b] OLS estimates.
[c] Two-stage, maximum likelihood estimates.

Table 2. Determinants of Lead Story Coverage

Variable	*Energy*[a]	*Unemployment*[b]	*Inflation*[b]
% Change in energy cpi	.67 (.18)		
Unemployment rate		.55 (.11)	
Change in unemployment rate (absolute value)		1.87 (.43)	
% Change in cpi (above mean only)			.72 (.40)
Presidential speech		.67 (.27)	1.92 (1.36)
Constant	1.27 (.82)	−3.25 (.73)	2.65 (.56)
Rho	.46 (.14)		
Adjusted R^2	.23	.66	.10
S.E. of Regression	2.46	.65	2.52
Durbin-Watson statistic	1.83	1.92	1.66

[a] Maximum likelihood estimates.
[b] OLS estimates.

brings forth four additional stories on that subject, while a change of slightly over .5 percent warrants a lead story.

We used the absolute value of the change in unemployment because we suspect that news coverage is driven by changes in economic conditions in either direction. The coefficient suggests that news coverage is indeed not limited to "bad news;" decreases in unemployment are as newsworthy as increases. We were unable to test for similar patterns in the areas of energy and inflation because there were no periods of deflation during the seven years we examined.

Inflation coverage, too, is driven by changing (worsening) national conditions, though here the changes must be unusually severe to warrant additional news attention. The percentage increase in the cpi must surpass the seven-year average (1.5 percent in two months) in order for it to affect news coverage. A one percent increase in the cpi, for example, generates no new stories. A two percent increase, however, produces between one and two lead stories and a total of about eight stories over the two-month period.[5]

---

[5] Other indicators of inflation had no effect. For total stories, changes in the food cpi, b = .84, standard error = 1.94; for lead stories, changes in the food cpi, b = .47, standard error = .43; for total stories, changes in interest rates, b = 1.12, standard error = .98; or lead stories, changes in interest rates, b = .38, standard error = 2.06.

It is not surprising that television news responds only to unusual surges in the inflation rate. The period under investigation was marked by continually high inflation (the average annual cpi was 8.9 percent). The networks thus may have taken inflation for granted, deeming the issue newsworthy only when it was unusually severe.

Presidential and other elite activity also affect news coverage. A presidential speech on energy generates nearly 10 stories, while OPEC meetings are followed by an average of seven additional energy stories. Meanwhile, speeches on both unemployment and inflation elicit lead story coverage, though in the case of the former the standard error of the coefficient is too high to allow much certainty.[6] Finally, both indicators of news coverage are as a rule immune to shifts in the level of public concern. In but one instance (total coverage of inflation) did we detect increased coverage following a rise in public concern. Even here, the effect is quite small—an increase of 3 percent in the level of public concern leads to but 2 stories in a two-month period.[7]

DISCUSSION

Our results demonstrate that in all three issue areas, television news coverage is at least partially determined by real-world conditions and events. We are impressed by this overlap between the real-world and news coverage, particularly in the case of unemployment, where nearly two-thirds of the variance in news coverage is explained by economic conditions. In general, our results suggest that television news coverage provides at least a partial reflection of the state of the nation.

News coverage is for the most part unaffected by public opinion and the assumption that agenda-setting is a recursive process is on solid ground.

---

[6] The probability that the true coefficient is greater than zero is approximately .8. Presidential speeches on energy do not produce lead stories (*b* = .52; standard error = .98). The same is true of OPEC meetings: *b* = .72, standard error = .62. The dates of speeches used in this analysis are listed in Appendix A.

[7] To calculate the effect of public concern, Fair's two-stage estimator was used. The level of public concern was treated as an endogenous variable, using as instruments presidential speeches on energy, the level of concern with unemployment, and the level of concern with inflation. These are the variables found in the second portion of the analysis to affect the level of concern for energy (see Table 3). The effect of concern for energy on the total number of stories was −.61, standard error = .80; the effect on the number of lead stories was .01, standard error = .08.

In the case of unemployment, the instruments used were the unemployment rate and the average duration of unemployment (see Table 3). The effect of public concern on total coverage was −.14, standard error = .17. The effect on the number of lead stories was .004, standard error = .03.

For inflation, presidential speeches, the importance of energy, and the importance of unemployment were used as instruments (see Table 3). The effect of concern for inflation on the number of lead stories = .03, standard error = .04.

Coverage of inflation provides the only instance of this "agenda-setting in reverse" effect. This effect may have resulted from the extraordinarily high level of public concern accorded inflation during the 1974–1980 period.

Not surprisingly, the president is a major contributor to the network agenda. Not only is a presidential address a newsworthy event in itself, it can also serve as a journalistic peg on which to hang a variety of feature stories (see Gans, 1980:168). That is, presidential activity causes a flurry of news coverage, not all of which concerns the original event. President Carter's 1979 televised address on energy, for instance, was followed by five separate stories that week on the Alaskan pipeline, problems facing the tourism industry, and the increase in reported thefts of gasoline.[8]

Determinants of the Public Agenda

We have found that real world conditions partially determine levels of media coverage. We now set out to investigate the impact of real world conditions and news coverage on the public agenda. This section of the analysis builds on our previous results. To correctly identify the two-stage model, we use as instruments those variables we earlier found to predict news coverage.

In two of three areas (energy and inflation), levels of public concern are indeed driven by television coverage (see Table 3). However, it is the *lead* story that serves as the agenda-setting force. The effect is substantial—every lead story dealing with energy raises the level of concern by about 1.25 percent, and each story related to inflation has a similar effect.[9] Only

[8] In general, the number of news stories reporting the occurrence and content of these presidential addresses represented only a fraction of the number of stories generated by such a speech. For example, an average of three stories was devoted by CBS to covering each speech on energy. This is less than a third of the average of 10 energy stories we find to be prompted by each speech. Thus, we have good reason to believe that a presidential address can generate stories not directly related to the speech itself.

[9] It is appropriate to treat the number of lead stories as an endogenous variable because the error terms of the equation used to predict its values are correlated with the error terms of the equation estimating its impact on the public's concern. However, because in the case of inflation we were unable to build a model accounting for much of the variance in the number of lead stories (see Table 2), the standard errors of the estimates for the effect of media on salience levels were quite large. The coefficients, themselves, however, were not significantly different from those estimated in the ordinary least squares regression. We therefore present the OLS estimates, with the caution that the coefficients for the media variables may be overestimated by 20 to 30 percent. This does not change our conclusion that these variables exert significant impact on the level of the public's concern with inflation.

Table 3. Determinants of Public Concern

Variable	*Energy*[a]	*Unemployment*[b]	*Inflation*[c]
Number of lead stories	1.25 (.22)	1.24	(.44)
Number of lead stories ($t-1$)			.82 (.47)
Unemployment rate		3.04 (1.36)	
Average duration of unemployment (weeks)		1.70 (.65)	
Presidential speech	5.77 (1.33)		6.07 (3.51)
Control variables			
Concern for energy			−.47 (.23)
Concern for unemployment	−.24 (.09)		−.62 (.24)
Concern for inflation	−.09 (.07)		
Constant	8.87 (4.17)	−30.34 (7.05)	44.84 (5.38)
Rho	.33 (.14)	.41 (.14)	
S.E. of regression	2.87	3.98	6.80
Adjusted R^2	.79	.46	.57
Durbin-Watson statistic		1.83	1.99

[a] Two-stage, maximum likelihood estimates.
[b] Maximum likelihood estimates.
[c] OLS estimates.

in the case of unemployment do we find no agenda-setting by lead stories.[10]

The number of nonlead stories has no independent effect on public concern for any of the three issues.[11] One should not necessarily conclude, however, that nonlead stories do not affect the public agenda. A more plausible interpretation is that the audience's attention span dissipates rapidly and that the impact of news stories diminishes with their placement in the newscast.

---

[10] For the simultaneous equations, maximum likelihood estimates, with media variables treated as endogenous, the number of lead stories had a coefficient of .31, with standard error = 1.03. Presidential speeches on unemployment, and the change in the rate of unemployment were used as instruments to identify the equations.

[11] For energy, b = .10, standard error = .15. The number of nonlead stories was treated as an endogenous variable, with presidential speeches and OPEC meetings serving as instrumental variables. For unemployment, b = .09, standard error = .35. The unemployment rate and the percentage of change in the rate (absolute value) were used as instruments. For inflation, b = .03, standard error = .19. The percentage of change in the cpi (above mean only) was used as the instrument.

Only in the case of inflation is there any lingering impact of television coverage. Here the number of stories during the previous time period as well as the number of stories in the current period determine the current level of concern.[12]

Public concern for issues responds not only to television coverage; real-world conditions and events have an independent impact in all three issue areas. For both energy and inflation, public concern is shaped directly by the president. For every presidential speech on the subject of energy, an additional 6 percent of the citizens' concerns relate to energy as a result. When the president speaks to the nation on inflation, the resulting rise is again about 6 percent. Only speeches on unemployment have no agenda-setting effect.

Unemployment is the only area where public concern is affected directly by economic conditions. Citizens' concern rises with the national level of unemployment and the average duration of unemployment. For every percentage point of the work force unemployed, the issue draws the attention of 3 percent of the public, while each week added to the average length of unemployment contributes 1.7 percent.[13] Note, however, that the impact of economic conditions has no long-term component. That is, citizens are unaffected by previous levels of unemployment when deciding whether the issue is currently an important problem.[14]

Notable for their failure to affect the salience of energy directly are economic indicators. Higher energy costs, higher fuel oil costs, energy consumption, and OPEC imports all fail to heighten directly public concern over energy.[15] Nor is public concern for inflation related to any indicator of economic conditions. Changes in the cpi, the cpi for food, and interest rates have no impact.[16]

---

[12] For energy, *b* = .05, standard error = .18.

[13] It is not surprising that public concern is related to the average duration of unemployment while news coverage of the issue is not. Duration is a relatively obscure statistic; the news media rarely cite it. On the other hand, the conditions that it represents are likely to be felt quite strongly by the people affected by unemployment. To them, the longer the period of unemployment, the more serious the problem.

[14] For the level of unemployment in the previous period, *b* = .31, standard error = .37. For the average duration of unemployment, *b* = .19, standard error = .14.

The control variables—concern for energy and inflation—have no effect in this equation. For energy salience: *b* = .08, standard error = .17; for inflation salience: *b* = .17, standard error = .13. Keeping the latter in the equation did not alter fit of the model significantly.

[15] Our results here are at odds with MacKuen's analysis, in which the price of home heating oil is found to directly affect the level of concern with energy (MacKuen and Coombs, 1981:94). In our analysis, for change in energy prices, *b* = .13, standard error = .27; for changes in heating oil prices, *b* = −.04, standard error = .42; for OPEC imports, *b* = .0001, standard error = .001.

While economic conditions exert no direct effect on citizens' concern for energy and inflation, there are mediated effects. Rising prices stimulate additional news coverage which, in turn, boosts public concern. A 6 percent increase in energy prices over two months, for example, is followed by four lead stories, which raise the level of public concern by just over 4 percent. The effect is similar in the case of inflation.

DISCUSSION

A summary of our results is presented in Figure 4. In two of the three issue areas we detect unambiguous evidence that television news sets the public agenda. Unemployment is the deviant case; here public concern is directed entirely by economic conditions. The distinctiveness of unemployment may stem from the low level of news coverage accorded the issue. CBS broadcast on average only one lead story on the issue every two months, compared to three leads for energy and four for inflation. This level of coverage may be insufficient to boost viewers' perceptions of the importance of the issue.[17]

The low level of coverage accorded unemployment by the networks suggests that information about unemployment is acquired from other sources. Given the severe personal consequences of unemployment, individuals may seek out information about unemployment in their own communities or occupations, information that is more apt to be conveyed interpersonally or by local media rather than network newscasts. In short, the sensitivity of public concern to national conditions may be mediated by local media or by a two-step flow of information.

Not only do we detect evidence of agenda-setting by television news, our results also indicate that the effect is in fact unidirectional. News coverage boosts public concern but public concern does not, for the most part, alter the level of coverage. The only instance of such a feedback effect involves coverage of inflation, a result that can be attributed to the extraordinarily widespread public concern for this issue during the time period under investigation, or similarly widespread concern among news executives.

---

[16] For percentage of change in cpi, *b* = .51, standard error = 1.76; for percentage of change in food cpi, *b* = .06, standard error = 1.18; for percentage of change in interest rates, *b* = −.37, standard error = .62.

[17] We should emphasize here that our data precede the dramatic increases in unemployment which occurred in 1981 and 1982. For the period under study, the rate of unemployment ranged from 5.0 to 8.9 percent. It is possible that, as in the case of inflation, a threshold effect governs coverage of unemployment. Thus when unemployment reaches an unusually high level, news coverage may rise substantially, possibly triggering an agenda-setting effect.

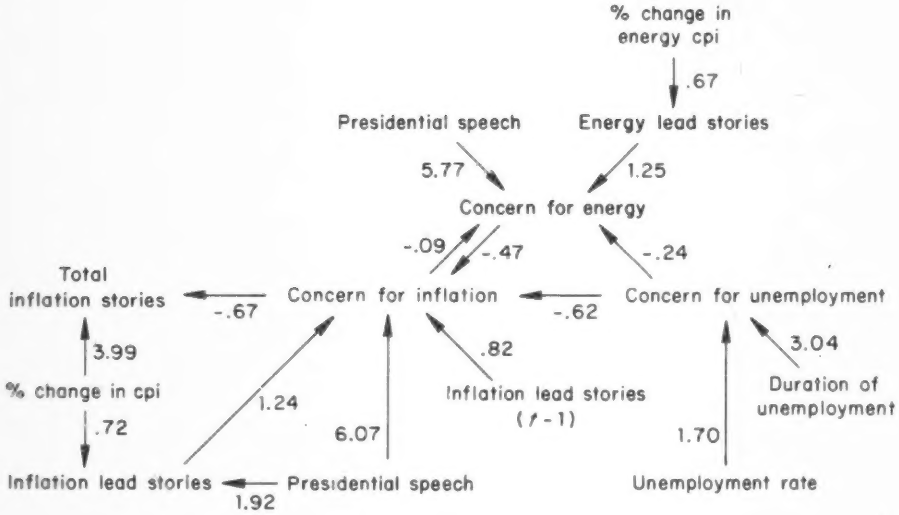

% change in energy cpi
.67
Presidential speech
Energy lead stories
5.77
1.25
Concern for energy
-.09
-.47
-.24
Total inflation stories
-.67
Concern for inflation
-.62
Concern for unemployment
3.99
.82
3.04
% change in cpi
1.24
Inflation lead stories (t - 1)
Duration of unemployment
.72
6.07
1.70
Inflation lead stories
Presidential speech
Unemployment rate
1.92

Figure 4. Summary of results

A major implication of our results is that news stories are not equally effective in molding the public agenda: the lead story in the newscast has the strongest impact on citizens' perceptions of issue importance. We suggest three explanations. First, from the networks' perspective, the lead story is the day's most important news (Gans, 1980). It is possible that the audience recognizes this distinction and therefore attaches greater weight to information conveyed in the form of lead stories. As noted earlier, the greater impact of lead stories over nonlead stories may also stem from a higher level of viewer attentiveness during the first few minutes of the newscast. Finally, research on attitude change suggests that the order in which persuasive messages are presented can affect the probability of message acceptance; arguments tend to be more persuasive when presented earlier rather than later (for a discussion of order-of-presentation effects, see Petty and Cacioppo, 1981). A similar "primacy" effect may underlie agenda-setting by television newscasts.

Our results also point to the president as an important contributor to the public's political agenda. When the president speaks, the public listens. And not only does the president exercise direct influence over citizens' concerns, but since presidential addresses are prominently positioned in the news, he has a mediated effect as well.

Finally, our results make clear that citizens' issue concerns are indeed volatile. Media and real-world effects are quickly supplanted. In but one instance were we able to detect an agenda-setting effect that persisted past two months. The short-term duration of these effects does *not* detract from their strength or reliability. Instead, these results suggest a public with a limited memory for past conditions and news stories, a public whose political concerns are heavily dependent on present conditions and news coverage.

Conclusion

The public agenda is indeed affected by what television journalists and editors choose to broadcast as news. However, news coverage is not the sole source of citizens' issue concerns. As real-world conditions worsen and the president turns his attention to particular issues, public concern for these issues increases. Real-world conditions and events provide an independent impetus to the perceived importance of issues. Moreover, since news coverage of issues is to a significant extent determined by actual conditions, analyses of media agenda-setting that ignore real-world conditions will arrive at severely inflated estimates of media influence.

It is true that the three issues examined here are relatively obtrusive—issues with tangible consequences for individuals. As such, our results may not be generalizable to issues that are more remote and symbolic. Public concern for symbolic issues may be a good deal more susceptible to media coverage than we found it to be for energy, unemployment, or inflation.

Not only do real-world conditions complement media as determinants of issue salience, but as Erbring and his colleagues have pointed out (1980), they may also govern individuals' sensitivity to the media agenda. Under conditions of rising unemployment, for example, citizens may become alert to the issue and thus monitor the media more vigilantly for news about unemployment. To the extent that an issue has personal significance, news about the problem can boost one's concern to even higher levels. In short, personal experience and media agendas may interactively shape citizens' political concerns (for some evidence of such an interaction effect, see Erbring, et al., 1980).

We searched unsuccessfully for interactions between real-world conditions and television news coverage. However, aggregate data such as ours are hardly appropriate for discerning interaction between *personal* experience and media coverage. Without the benefit of individual-level analysis, it is impossible to test the interaction hypothesis.

While our results suggest that real-world conditions trigger agenda-setting by television, we can only speculate over the processes by which real-world conditions affect public opinion. Three possibilities may be considered. First, public concern for issues may vary directly in response to national conditions. As energy shortages worsen, more drivers spend time in gas lines; as food prices rise, more shoppers notice the decline in their purchasing power. Second, worsening national conditions spawn news coverage on the basis of which people form judgments about national problems. As inflation increases and the networks run more stories on inflation, more people conclude that inflation is an important national problem. Finally, people may respond to information about local conditions in their neighborhoods or communities, information acquired through interpersonal communication, direct experience, or local media. Local information may be generalized to form judgments about national conditions. Monitoring these intervening processes is vital if we are to understand the dynamics of the public agenda.

Appendix A

SOURCES AND CALCULATION OF MOST IMPORTANT NATIONAL PROBLEM FIGURES

In order to estimate the extent to which determinants of the public agenda have a lasting effect, it is necessary to have observations taken at equal intervals. No single survey organization asks the question "what do you think is the most important national problem facing the nation?" with any regularity, however. By using the results of three polls—AIPO, Yankelovich, and University of Michigan—we were able to obtain a reading for every two-month period between 1974 and 1980, save three. In their original state, however, results from different organizations were not compatible. Both AIPO and Yankelovich accept multiple answers, while Michigan does not. Additionally, Yankelovich consistently records a higher average number of answers—sometimes over two per respondent. Gallup generally averages under 1.5 mentions per person. Thus, if one compares readings of two or more organizations in any of the months where their polls are conducted simultaneously, as a rule, Yankelovich will record higher percentages of *respondents* citing any given issue. Michigan polls record the lowest figures.

To make the three sets of results compatible, the percentage of *responses* concerning each issue was calculated. This transformation yielded essentially equal readings for the several months in which more than one organization asked the question. For example, in August 1979, both AIPO and Yankelovich asked about the most important national problem. Of AIPO respondents, 19.7 percent said energy, 7.5 percent cited unemployment, and 50.3 percent mentioned inflation. In contrast, of Yankelovich respondents, 35.4 percent listed energy, 10.1 percent named unemployment, and 81.0 percent said inflation. Clearly these figures are incompatible. However, when the percentage of *responses* was calculated, AIPO had 15.4 percent for energy, 5.8 percent for unemployment, and 39.3 percent for inflation. Of Yankelovich's responses, 18.3 percent dealt with energy, 5.2 percent with unemployment, and 43.8 percent with inflation. Here, the differences are statistically insignificant. On the basis of several such comparisons, we concluded that calculating the percentage of *responses* allowed us to use three different polls interchangeably.

In the several two-month periods when the "most important national problem" question was asked more than once, either by the same or by different organizations, the respondents were pooled, and then the percentage of responses expressing concern with each issue area was calculated. In the three cases where no observations were available, levels were estimated by interpolation.

The following surveys were used: AIPO 886, 906, 913, 915, 916, 924, 932, 938, 943, 950, 960, 961, 970, 980, 986, 993, 999, 1106, 1111, 1123, 1128, 1136, 1141, 1151, 1157, 1159, 1162; Yankelovich 8400, 8422, 8430, 8440, 8460, 8510, 8520, 8530, 8550, 8105, 8117, 8125, 8149, 8156, 8161, 8181, 8182, 8184, 8260; University of Michigan 1974, 1976, 1978.

Appendix B

The equations comprising the model can be written as follows:

$$\text{TV News} = c + \text{Real World} + \text{Issue Importance} + e \qquad [1]$$
$$\text{Issue Importance} = C + \text{TV News} + \text{Real World} + E \qquad [2]$$

Equations 1 and 2 represent the "theoretical" model. Because of the transformations required to make the data from the three survey organizations compatible, responses citing one issue necessarily cause the percentage of responses referring to other issues to decrease (see Appendix A). Thus for example, the mentioning of energy as an important problem causes a decrease in the percentage of responses related to unemployment. To adjust for this statistical relationship, each equation must include the percentage of responses citing the two other issues. Thus the operational version of the model to be estimated for each issue is as shown in Figure 5.

It is appropriate to estimate the above simultaneous equations model with a two-stage procedure. Failure to use a two-stage estimator would produce inconsistent (upwardly biased, in this case) coefficients. This is because the determinants of the "dependent" variable omitted from the model may be correlated with omitted determinants of the "endogenous" variables (that is, the error terms of the equations may be correlated). In addition, a two-stage estimator makes it possible to investigate the direction of causality. One of our goals in this endeavor is to determine whether media agendas affect public agendas, and vice-versa (for illustrative applications of two-stage analysis, see Jackson, 1975; Jacobsen, 1975; Markus and Converse, 1979).

A key requirement of two-stage estimation is the specification of exogenous variables. In order to estimate the model pictured above, we must identify (1) causes of news coverage that exert no direct impact on public concern (this allows us to estimate the effect of media attention on problem importance), and (2) variables that directly influence public concern, but not news coverage (to allow us to estimate the feedback effect).

Ideally, the selection of instruments is guided by a priori theoretical considerations. But theories of "newsworthiness" are remarkably general. For example, newsworthy events are said to be those that are salient to the audience, reflect changing conditions, involve important decision makers or large numbers of people, or are current. By these criteria, few events are not newsworthy. We are thus forced to rely on intuitive expectations about which events and conditions warrant news coverage. Our strategy is to examine a series of

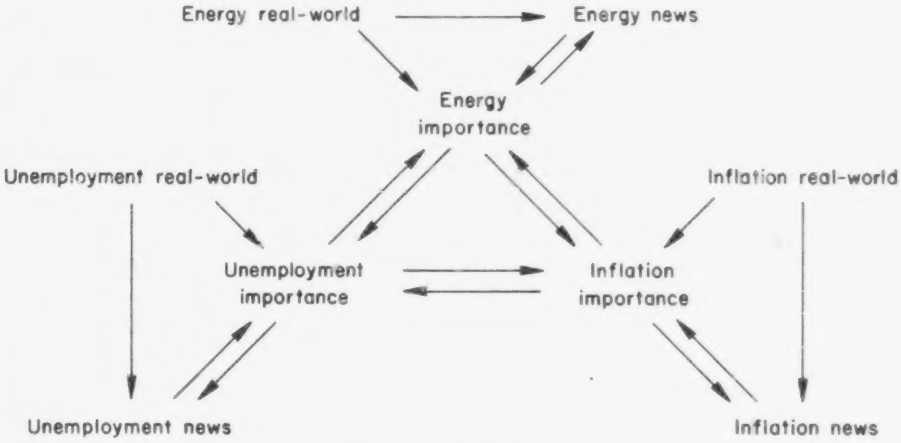


Figure 5. Model to be estimated

plausible determinants of news attention. Those which we find to be related to news attention we retain; the remainder we discard. We follow a similar course to find determinants of public concern.

We fully recognize the post-hoc nature of our instrument selection and model specification strategy. But, since there is so little hard evidence on the interplay between real-world conditions, news coverage, and public concern, we feel that it is preferable to estimate a two-stage simultaneous model in this manner than to fall back on the conventional assumptions of recursiveness and nonsimultaneity. Furthermore, there is good reason to believe that our instruments are indeed true causes of public concern and news coverage.

In order to estimate properly each of the *k* equations in the model ($k = 6$), there must be at least *k*-1 variables excluded from each equation. We meet this requirement; all our equations are sufficiently identified. In addition, the efficiency of the coefficients estimated at the second stage is in part a function of the success of the first stage estimates. Our ability to make these first stage estimates varies from one equation to another. So too, then, does the efficiency of our second stage estimates.

In addition to the difficulties of simultaneity, we must also address the possibility that the error terms in the equations may be autocorrelated across time. Positive autocorrelation causes one to underestimate the size of the standard errors and hence leads to overconfidence in the estimates.

When there is both simultaneity and autocorrelation, the simultaneous equations, maximum likelihood estimator developed by Fair (1970) is used. The estimator works in the following manner: first, values of the right-hand side endogenous variables (public concern in the first part of the analysis, television coverage in the second) are estimated, using as instruments (1) lagged values of the left-hand side endogenous variable, (2) lagged values of the right-hand side endogenous variables, (3) current and lagged values of the exogenous causes of the left-hand side endogenous variables (real-world conditions and events), and (4) current values of the exogenous causes of the right-hand side endogenous variables. Then the effects of the right-hand side endogenous and exogenous variables on the left-hand side endogenous variable are estimated using an iterative search procedure to determine, and correct for, rho, which measures the extent of the autocorrelation. (See Fair, 1970:508–9.)

When only autocorrelation is a problem, a simple maximum likelihood estimator is used (see Johnston, 1972). When only simultaneity exists, we use two-stage least squares.

References

Altheide, David, L.
1976 Creating Reality: How Television News Distorts Events. Beverly Hills: Sage Publications.

Behr, Roy, L.
1984 "The effects of media on voters' considerations in presidential and congressional elections." Unpublished paper, Department of Political Science, Yale University.

Benton, M., and P. J. Frazier
1976 "The agenda-setting function of the mass media at three levels of information holding." Communication Research 3:261–74.

Bower, Robert, T.
1983 "Trends in media attitudes." Paper delivered at the Annual Meeting of the American Association for Public Opinion Research.

Epstein, M. J.
1973 News from Nowhere. New York: Random House.

Erbring, Lutz, E. N. Goldenberg, and A. H. Miller
1980 "Front-page news and real-world cues: a new look at agenda-setting by the media." American Journal of Political Science 24:16–49.

Fair, Ray C.
1970 "The estimation of simultaneous equation models with lagged endogenous variables and first order serial autocorrelated errors." Econometrica 38:507–16.

Gans, Herbert J.
1980 Deciding What's News. New York: Vintage Books.

Iyengar, Shanto, D. R. Kinder, J. Krosnick, and M. D. Peters
1984 "The evening news and presidential evaluations." Journal of Personality and Social Psychology 46:778–87.

Iyengar, Shanto, M. D. Peters, and D. R. Kinder
1982 "Experimental demonstrations of the "not-so-miminal" consequences of television news." American Political Science Review 76:848–58.

Iyengar, Shanto, and D. R. Kinder
In Press "More than meets the eye: television news and public evaluations of the president." In G. Comstock (ed.), Advances in Public Communication: Volume 1. New York: Academic Press.

Jackson, John E.
1975 "Issues, party choices, and presidential votes." American Journal of Political Science 19:161–86.

Jacobson, Gary C.
1980 Money in Congressional Elections. New Haven: Yale University Press.

Johnston, J.
1972 Econometric Methods. New York: McGraw-Hill

Lippman, Walter
1922 Public Opinion. New York: Harcourt Brace.

MacKuen, Michael J., and S. L. Coombs
1981 More Than News. Beverly Hills: Sage Publications.

Markus, G. B., and P. E. Converse
1979 "A dynamic, simultaneous equations model of electoral choice." American Political Science Review 73:1055–70.

McCombs, M.
1976 "Newspapers vs. television: mass communication effects across time." In M. McCombs and D. Shaw (eds.), The Emergence of Political Issues. Denver: West Publishing Co.

Meadow, R. G.
1973 Cross-media comparison of coverage of the 1972 presidential campaign. Journalism Quarterly 50:482–88.

Patterson, Thomas E., and Robert D. McClure
1976 The Unseeing Eye. New York: Thomas Crowell.

Petty, Richard E., and J. T. Cacioppo
1981 Attitudes and Persuasion: Classic and Contemporary Approaches. Dubuque, Iowa: Wm. Brown Co.

Public Opinion Magazine
1981 "Face off: a conversation with the president's pollster." March, pp. 2–12.

A Critical Analysis of Australian Television Coverage of Election Opinion Polls

TED J. SMITH III AND DEREK O. VERRALL

SUCCESSFUL opinion polling requires the active cooperation of the general public. This cooperation is likely, however, only if the polls are believed to provide accurate readings of public opinion. Failing this, there is little incentive for anyone to either participate in opinion surveys or provide them with honest and thoughtful information. Whether polls are accurate is the responsibility of the pollsters; whether they are perceived to be accurate is largely a function of their coverage in the news media.

The link between polling and the press has been the focus of extensive recent discussion, including a 1980 symposium in *Public Opinion Quarterly* and a 1979 conference sponsored by the National Council on Public Polls (NCPP) and the Kettering Foundation (see Cantril, 1980, for proceedings). Critics from the United States (e.g., Germond, 1980; Ladd, 1980; Wheeler, 1976, 1980), the United Kingdom (e.g., Barter, 1982; Crewe, 1982; Worcester, 1980), Australia (e.g., Pryor, 1980; Thompson, 1981), and West Germany (Noelle-Neumann, 1980) have

Abstract Much recent attention has focused on the nature and possible effects of media coverage of opinion polls. Accordingly, this article reports the results of a detailed critical analysis of Victorian television news coverage of election opinion polls during the 1980 Australian federal election campaign. Five aspects of the coverage were analyzed: its extent, focus, completeness, accuracy, and critical balance. The results of the study indicate that the coverage was seriously, perhaps irredeemably, flawed.

Ted J. Smith III is an Assistant Professor of Rhetoric and Communication Studies at the University of Virginia. Derek O. Verrall is a Lecturer in Politics at the Warrnambool Institute of Advanced Education. This research was conducted with the assistance of a Warrnambool Institute Faculty Research Grant. The authors would like to thank the following individuals for their invaluable contributions to the project: Ray Bantow, Barry Bron, Tony Daggett, Tony Loquet, Bruce Pettingill, Rosemary Smith, and Ian Ward.

Public Opinion Quarterly Vol. 49:58–79 © by the Trustees of Columbia University
Published by Elsevier Science Publishing Co., Inc. 0033-362X/85/0049-58/$2.50

been unanimous in the claim that reporting of poll results is at best superficial and at worst inaccurate and misleading.

Although anecdotal and impressionistic evidence for this claim abounds, very few systematic studies of poll coverage have been published. The major exceptions are three recent American studies. Paletz, et al. (1980) examined all poll stories in the *New York Times* and CBS and NBC weekday evening television news programs for the years 1973, 1975 and 1977. Broh (1980) examined poll stories in the *New York Times, Time, Newsweek* and national television news for the last two months of the 1976 presidential campaign. Finally, Miller and Hurd (1982) examined a sample of stories drawn from three major newspapers over the period 1972–79 in order to assess compliance with the American Association for Public Opinion Research (AAPOR) reporting standards. Taken together, these studies support three very tentative conclusions.[1] First, media coverage of opinion polls was extensive, especially during election campaigns, and increased dramatically throughout the period of study (1972–79). Second, this coverage was superficial in the sense that basic methodological details were frequently omitted. This problem was particularly serious in the case of television reports, but less serious in the case of newspaper accounts of election surveys. Third, fragmentary evidence from Paletz et al. (1980:507) and Broh (1980:517) suggests that media coverage was often inaccurate.

If these conclusions are valid, they raise issues of concern to both pollsters and journalists. Further research, focused in particular on the question of media accuracy, is clearly warranted. Accordingly, this article reports the findings of a detailed analysis of television news coverage of election opinion polls published during the 1980 Australian federal election campaign. Five characteristics of this coverage are examined: its extent, focus, completeness, accuracy, and balance. The exclusive focus on television coverage can be justified on the grounds that in Australia, as in many other industrialized countries, television has emerged as the most popular and respected source of news and information (FACTS, 1980:24). The exclusive focus on coverage of election polls is justified by the often repeated claim (e.g., Ladd and Ferree, 1981:14) that the polling industry is judged primarily by its performance on these polls.

The remainder of the article is divided into four sections.The first provides background information about the 1980 Australian federal elections. The other three sections outline the methods used in the study, report its results, and discuss their significance.

---

[1] Media coverage may have changed over the past few years. Given the existence of powerful structural constraints on poll reporting, this seems unlikely. But see McKee (1982) and Noelle-Neumann (1980) for examples of innovative coverage.

The 1980 Australian Federal Elections

Australia is a parliamentary democracy. Its legislature consists of the House of Representatives (125 seats in 1980) and the Senate (64 seats). After an election, whatever party or coalition holds a majority of seats in the lower house has the right to form a Government. Thus in Australian elections attention is focused primarily on the contest for the House of Representatives.

There are four major political parties in Australia: Liberal Party of Australia (LP), National Country Party of Australia (NCP), Australian Labor Party (ALP), and Australian Democrats (AD). The first two parties have traditionally operated as a coalition (L/NCP) at the national level. The Australian Democrats, while represented in the Senate, have never won a seat in the House. Therefore, in practice, House of Representatives elections reduce to a two-way struggle between the L/NCP coalition and the ALP.

Elections to the House of Representatives must be held at least every three years, but it is the prerogative of the prime minister to call an election at some earlier date of his choice. Australian election campaigns are quite brief, usually lasting no more than six or seven weeks. The process begins when the prime minister sets a date for the election. Each political party then spends several weeks developing an election platform. Obviously, campaign activities are held to a minimum during this period. Each party's platform is then communicated to the voters by means of a formal policy speech. The date when the first of these speeches is delivered marks the beginning of the "official" campaign, which lasts for approximately three weeks and culminates in the election.

Three features of the Australian political system are of particular interest. First, voting is compulsory for all citizens over the age of 18. This is a mixed blessing for those who attempt to predict election results. On the one hand, it contributes to the highest voter turnout rate in the world, an average of 95.4 percent in the 14 elections between 1945 and 1980 (Crewe, 1981:53). On the other hand, it insures that many undecided or indifferent voters will, in fact, vote. Second, in House of Representatives elections voters are required to indicate their order of preference for all candidates standing in the electorate.[2] If no candidate receives a majority of first preference votes, then preferences are distributed beginning with the least popular candidate. This process continues until one candidate receives an absolute majority. This system of voting adds many complexities to Australian political discussions. Not least of these is the fact that support for a candidate can be expressed (in an election or an election opinion poll) in two very different ways. The *first preference* vote simply reflects the percentage of voters who give their first preference to a particular candidate. The *two-party preferred* vote treats every seat as a contest between the L/NCP coalition and the ALP. This vote is computed by adding the first preferences for the two coalition partners and allocating the votes for the minor parties (including the Australian Democrats) to the coalition or the ALP according to expressed preferences (in the case of elections) or some formula (in the case of opinion polls). This means that there are also two methods of computing "swing": first preference and two-party preferred. Although it remains controversial, media discussions are conducted almost invariably in terms of the two-party preferred swing thought necessary for the government to change hands.[3] The third interesting feature is the existence of a "blackout" law. Elections are always held on a Saturday. Beginning at midnight on the Wednesday preceding the election, the broadcast media are prohibited from carrying any election-related material. Thus voters are forced to rely exclusively on the print media for political information during the last few days of the campaign. In practice, this means that campaign activities begin to slow down on Wednesday evening. The law also has a profound effect on the conduct of election opinion polls.

---

[2] Senate elections use a totally different and highly complex method of proportional representation. See Mackerras (1980:333–35) for a brief discussion.

On December 10, 1977, the L/NCP coalition was returned to power with a majority of 48 seats, the second largest majority in Australian electoral history. The coalition received an estimated 54.6 percent of the two-party preferred vote. On September 12, 1980, Liberal Prime Minister Malcolm Fraser announced that new elections would be held on October 18. The first policy speech was delivered on September 30; the "blackout" on electronic news came into effect at midnight on October 15. Thus the official campaign in 1980 was 19 days long, 16 for the electronic media.

Early in September the consensus of informed observers (and the general public) was that the coalition would be returned to power with a substantial but reduced majority—something on the order of 25 seats—after a tedious and predictable campaign. They were partially correct. The Government was returned with a substantial but reduced majority of 23 seats. The coalition parties received an estimated 50.4 percent of the two-party preferred vote, a 4.2 percent swing to Labor. The campaign, however, was anything but tedious. This was due in large part to the findings of the election opinion polls.

---

[3] The major argument against the use of two-party preferred swing is that it is based on two estimated parameters, the actual two-party preferred vote in the preceding election and the percentage of the vote required for the ALP to gain power. Despite its shortcomings, the two-party preferred method of computing swing seems to be as accurate as any other indicator in predicting the number of seats that will actually change hands in a given election (Snider, 1976).

There are four major opinion polls in Australia: the Morgan Gallup Poll produced by the Roy Morgan Research Centre, the Australian Public Opinion Polls-The Gallup Method (APOP) produced by McNair Anderson Associates, the Australian Nationwide Opinion Poll (ANOP) produced by an organization of the same name, and the *Herald* Survey/*Age* Poll produced by Irving Saulwick and Associates.[4] These organizations conducted 10 separate surveys of voting intentions during the 1980 campaign. Their surveys shared three important similarities: all were commissioned by daily or weekly news publications, all were published during the period of the official campaign, and, most important, all showed a substantial swing to the ALP. In the end, three of the four final polls (APOP, ANOP, *Herald/Age*) "predicted" an ALP upset; the fourth (Morgan) "predicted" that the coalition would be reelected by a narrow margin (3–13 seats).

In the event, three of the four polls were "wrong" and none of them provided an accurate indication of the size of the coalition's majority. The significance of this result is discussed in the final section below. The prior and more important question, however, concerns the coverage of the polls by the mass media, especially television. Given a situation in which the poll findings clearly contradicted expert and popular expectations, how were those findings presented to the public? The study reported here was designed to answer that question.

Methods

The data base for study included all early evening television news programs broadcast in the state of Victoria during the official 1980 election campaign, that is, from September 30 (the date of the first policy speech) through October 15 (the date of the broadcast blackout). The data base was limited to a single state because of the nature of the Australian television system. There are no American-style television networks in Australia, although some material is shared among commonly owned stations. As a result, all television newscasts are produced by individual stations. These stations are concentrated overwhelmingly in the capital cities of the six states and two federal territories that make up the country. Given the wide geographic dispersion of originating stations (Australia is the size of the continental United States), and the lack of any television news archives, the study of television news content on a national basis is impractical. However, the concentration of stations in state capitals facilitates study of television news within a given state.[5]

---

[4] A new fifth organization, Spectrum, also conducted at least two telephone surveys during the campaign. These were commissioned by the national daily newspaper *The Australian*, and results were reported there in fragmentary and sometimes contradictory terms. It even appears that the same poll results were reported in issues a week apart as ostensibly different polls. Perhaps for these reasons, the Spectrum polls were ignored by other news media: no clear reference was made to them on Victorian television. Hence they were exlcuded from this discussion.

Melbourne, the second largest city in Australia, is the capital of the state of Victoria, the second most populous state in the country. In 1980, there were four television stations in Melbourne, three commercial and one public. Each of these produced a major daily newscast, starting between 6 and 7 P.M. and covering international, national, state, and local (i.e., Melbourne) affairs. These newscasts were rebroadcast throughout the state via affiliates and repeater stations and thus provided all television coverage of national affairs in Victoria. This study analyzed the output of the four stations for the 16 days of the official campaign.

The research reported here was conducted as part of a much more comprehensive study of Victorian print and broadcast coverage of the entire 1980 campaign. In the course of the larger project, videotape recordings were made of all evening television newscasts and all news items relating to the election were identified. For the present study, all of the election news items were examined for explicit references to election opinion polls. Any item containing one or more such references was designated a "poll item" and transcribed in its entirety. These transcriptions (and any relevant video material) were then analyzed on a word-by-word basis in order to provide at least partial answers to the following questions:

1. What was the *extent* of television coverage of opinion polls, in terms of both amount and emphasis?
2. Did television coverage *focus* primarily on the results of the polls, their meaning, or their effects?
3. Did television provide *complete* information about poll results as mandated by Australian Press Council guidelines?
4. Was television poll coverage generally *accurate* or did it contain clear errors of fact or interpretation?
5. Did television provide a *balanced* account of the strengths and weaknesses of election opinion polls?

Data coding followed a four-step process. First, relevant coding categories and associated recording instructions were developed. These are described in the next section. A sample of poll stories drawn from the period immediately preceding the period of study was then coded and categories and instructions were revised as needed. In the third step, each of the authors coded all of the poll stories from the official campaign using the revised categories and instructions. Finally, the two sets of data generated in the third step were compared and all discrepancies noted. Each discrepancy was then discussed and a final coding decision was made by consensus.

---

[5] The question arises as to the quality of Victorian television news compared to the other states. No definitive evidence is available, but given the high level of competition, Victoria's reputation for sophistication, and the subjective impressions of the authors, it seems likely that Victorian news is at least as good as the majority of other states.

Coding of data by the researchers, the procedure used in this study, has been attacked by Krippendorff (1980:74) as "probably the worst practice in content analysis" on the grounds that it prevents independent reliability checks and produces unreplicable results. He suggests using "independent" coders who are trained by working through a standard sample of material, comparing their results at each step with "ostensibly correct results" as determined by a "panel of experts" (the researchers?). Individuals who are not "suitable" for the task are then eliminated. Although Krippendorff's concern for reliability and replicability is highly laudable, it is difficult to see how the reliability checks generated by this or any other extensive training procedure are "independent" in any meaningful sense of the word. Further, to the extent that trained coders are less "expert" than the researchers who trained them, validity is decreased. Finally, successful replication depends less on who codes the material than on the clarity and completeness of the recording instructions. This study was designed in light of these considerations. Every effort was made to insure that the results would be valid, reliable and replicable.[6] In deference to established usage, however, it is perhaps better described as a critical analysis than as a "scientific" content analysis.

Results

This section reports findings relevant to the five questions noted above. The procedure followed in each case is to outline the nature and derivation of the coding system used and note the most salient results.

EXTENT OF COVERAGE

The first goal of the study was to determine the amount of coverage given to election polls. The superordinate category used in this analysis is the news *item*. News items are of two types: headlines and stories. A *headline* is a brief statement of news content given at the beginning of a news program (lead headline), within the program generally before a commercial break (internal headline), or at the end of the program (final

---

[6] Copies of transcripts and recording instructions are available from the authors.

headline). All four of the television stations made at least occasional use of all three headlines. A *story* is defined as an internally coherent segment of content. An item was coded as a story if it was initiated by the anchorperson, proceeded through one or more reporters, and returned to the anchor, or if it was presented exclusively by the anchor and was differentiated from other material by any of a range of verbal and nonverbal cues.

Two more specific types of news items are used in the study. An *election item* is a story or headline that makes explicit reference to the election campaign.[7] A *poll item* is an election story or election headline that makes explicit reference to an election opinion poll. Thus poll items are a subset of the larger category of election items.

An earlier study (Smith, 1981b) examined the amount of coverage devoted to the election as a whole. It found that the four television stations presented a total of 205 election stories and 80 election headlines during the 16 days of the official campaign (1981b:16). The election stories consumed 341 minutes of broadcast time, approximately 15 percent of total program time for that period (1981b:17).

The present study examined the amount of election coverage devoted to opinion polls, and found that the four stations presented 79 poll stories and 13 poll headlines during the campaign period. Thus 38.5 percent of all election stories, and 32.3 percent of all election items, contained at least one explicit reference to election opinion polls. It could be objected, however, that these figures overestimate the amount of poll coverage. For example, an election story that makes only a passing reference to opinion polls would be counted as a poll story. Therefore each poll story was further categorized according to the amount of content devoted to opinion polls. Three levels of content were distinguished: a poll *mention* (less than 10 percent of the story devoted to polls), a poll *theme* (10–50 percent), and a poll *focus* (more than 50 percent). Of the 79 poll stories 27 were poll focus stories; an additional 47 were poll theme stories. Thus nearly all of the poll stories (93.7 percent) contained substantial discussion of poll material.

These results provide strong support for the claim that opinion polls were a dominant focus in Victorian television coverage of the 1980 campaign. This conclusion is reinforced by two additional findings. The first relates to election headlines. Headlines on television news reflect journalistic judgments as to the relative importance of events on a given day. Thirteen (16 percent) of the 80 election headlines made explicit reference to opinion polls. But this figure is somewhat misleading. Poll headlines, by definition, make reference to specific events. The majority of election headlines, however, were general in nature (e.g., ". . . the latest news from the election trail . . ."). Thus the 13 poll headlines constituted 35 percent of the 37 election headlines which made reference to specific events; 9 of the 13 were lead headlines. Second, the 205 election stories included a total of 189 direct statements by political leaders. Of these, 37 (20 percent) made explicit reference to polls or were in response to a question which made such a reference (Smith, 1981b:20).

---

[7] See Smith (1981a, 1981b) for detailed definitions of the various election items.

FOCUS OF COVERAGE

The second goal was to determine, at least in a primitive way, what was said about the election polls on television news. This required an analysis of the general content of the 92 poll items. The typology used, and corresponding results, are best described in stages.

At the most basic level, information provided about an opinion poll can address one of only three issues: the findings or results of the poll, the meaning and significance of those results, or the actual or supposed effects of the poll on individuals or social groups. Accordingly, all of the verbal material relevant to polls in the 92 poll items was categorized under three headings: *results*, *meaning*, or *effects*. This allowed computation of the number of poll items in which each type of information appeared. As shown in Table 1, coverage was not restricted to mere reporting of poll results. Instead, all three issues were discussed extensively, but with a relatively greater focus on results and effects.[8]

The second level of the typology concerns coverage of each of the three issues. Thus coverage of results was categorized as *specific* (at least one quantified datum reported) or *general* (no quantified data). Although 70 (76 percent) of the 92 poll items reported poll results, only 15 items (16 percent) included any quantified data. This means the discussion of poll results on television news was conducted largely in general terms. It should also be noted that poll results were often "pluralized": the viewer was told, in general terms, what the poll*s* had found; reports of individual poll results, general or specific, were relatively rare.

Coverage of the meaning of poll results was also divided into two categories. Assertions and speculations about the factors that had produced the situation reported by the polls were categorized under the heading of *causes*. Inferences drawn from the poll results, and all discussions bearing on those inferences, were categorized under the heading of *implications*. Implications were further divided into *past*, *present*, or *future* on the basis of verb tense.[9] Coverage focused overwhelmingly on present and future implications. Only eight stories discussed possible causes; only three reported past implications. This is hardly a surprising discovery. Crespi (1980:466) has noted that "news is what is happening *now*, or is about to happen, not what happened last week—or even yesterday." But it is equally important to note that the election polls discussed in these reports actually "happened last week." Eight of the ten election polls were released before the broadcast blackout. The time lag between the start of interviewing and initial publication of these polls averaged eight days.

---

[8] This pattern is particularly clear in the case of poll headlines. Of the 13 headlines, 8 reported results, 1 examined meaning, and 8 asserted effects.

[9] To be coded as a future implication, a statement had to include an explicit transposition of results or their implications into the future (e.g. "Today's poll shows the election will be

Finally, claims about the existence of poll effects were divided into two categories: explicitly *stated* (e.g., "The latest polls have produced a new mood of optimism among ALP supporters.") and *implied* (e.g., "The polls still show Labor well ahead of the government. A new mood of optimism can be felt in the ALP camp."). Explicit discussion of possible effects was the most frequent category of poll coverage, appearing in 75 percent of all poll items (Table 1). Thus, the issue of poll effects emerged as the dominant organizational and interpretive focus of television coverage of the campaign. In general, three types of effects were attributed to the polls: changes in the mood of politicians, party supporters, journalists, and the general public; revisions of campaign strategy and tactics; and movements in share prices on the stock markets. The last of these developed into a major campaign issue. Journalists and coalition spokesmen attributed two dramatic drops in share prices to poll-induced fears of

Table 1. Content of Poll Coverage

Type of Focus	*Number of Items*[a]	*% of All Poll Items (n = 92)*[a]
Results	70	76%
Specific	15	16
General	65	71
Meaning	52	57
Causes	8	9
Implications	50	54
Past	3	3
Present	29	32
Future	36	39
Effects	77	84
Stated	69	75
Implied	34	37

[a] Subcategory totals exceed category figures because a given item can contain more than one type of content.

---

close"). Any doubtful case was coded as a present implication. Thus, present implications formed a residual category in a restricted sense.

a Labor victory. Labor spokesmen vehemently denied this interpretation and pointed to other possible causes. This exchange highlights a potentially serious problem with the poll coverage. Although the link between poll results and alleged effects was usually plausible, evidence of this link was seldom presented. Effects were simply asserted by journalists, often without regard to plausible alternative or contributory causes. Further, in some cases (e.g., widespread mood changes, some shifts in strategy or tactics), it is not at all clear that any "effect" did in fact occur.

INFORMATION PROVIDED

The third goal of the study was to determine what information was provided in reports of poll results. It is a truism that the findings of an opinion poll are uninterpretable in the absence of such methodological details as population, interview date, and question wording. Thus, various professional associations have promulgated guidelines for the minimum information to be provided in media reports of poll results. In the United States, standards of disclosure have been proposed by the American Association for Public Opinion Research (AAPOR) and the National Council on Public Polls (NCPP). In Australia, similar guidelines have been developed by the Australian Press Council (APC). All these are summarized in Table 2.

With these guidelines, it is possible to *evaluate* the adequacy of television coverage of poll results, but they stipulate only what information *ought* to be provided; they do not exhaust the range of information that *can* be provided. Therefore an attempt was made to expand the guidelines into a comprehensive list of all types of information relevant to poll results. This list, shown in Table 2, makes it possible to exhaustively *describe* the information provided in television coverage of poll results.

All material coded under the heading of "results" was examined in this phase of the study. The analytical procedures used were dictated by the nature of television poll coverage. Unlike newspapers, television news often describes the results of two or more polls in the same news item. Thus the appropriate unit of analysis for studying television coverage of poll results is the news *report*, defined as a self-contained description of the results of a single poll or summary statement of the results of multiple polls. A report can be specific (include quantified data) or general (no quantified data). Any particular poll item can include a number of reports of poll results. For example, three polls were released on October 15, the day of the broadcast blackout. A typical television poll story on that day began with a summary statement by the anchor of the results of the three polls (coded as one general report) and then moved to a reporter who described the results of each poll in some detail (three specific reports).

As noted above, 70 items covered poll results. These included a total of

Table 2. Information Provided in Reports of Poll Results

APC	*AAPOR*	*NCPP*	*Datum*	*Specific Reports (n=21)*		*General Reports (n=95)*		*All Reports (n=116)*	
				n	*%*	*n*	*%*	*n*	*%*
X	X	X	Sponsor: party for whom the poll was conducted	1	5%	1	1%	2	2%
X	X	X	Question wording: complete wording of question(s) asked in the poll	0	0	0	0	0	0
X	X	X	Population: definition of population from which sample was drawn	1	5	2	2	3	3
X	X	X	Sample size: total number of persons interviewed in the poll	0	0	0	0	0	0
X	X	X	Interview method: e.g., in person, mail, etc.	0	0	0	0	0	0
X	X	X	Timing: date(s) interviews were conducted	9	43	8	8	17	15
X	X	X	Partial results: identification of and basis for results that use less than the total sample	0[a]	0	–	–	–	–
X			Interviewers: who conducted the interviews, e.g., trained interviewers, journalists, etc.	0	0	0	0	0	0
X			Sampling method: e.g., random, stratified	0	0	0	0	0	0
X			Pollster: organization that conducted poll	10	48	2	2	12	10
	X		Sampling error: estimated percentage of error at some level of probability	0	0	0	0	0	0
			Response options: range of possible answers available to respondents	0	0	0	0	0	0
			Complete data: all first or two-party preference figures given	10	48	–	–	–	–
			Identification of data: data identified as first or two-party preference	6	29	–	–	–	–
			Undecided data: % undecided given	0	0	–	–	–	–
			Assumptions: basis given for distribution of preferences, allocation of undecided	0	0	0	0	0	0
			Place of publication: publication in which results first appeared	10	48	12	13	22	19
			Release date: date of first publication of results	12	57	20	21	32	28
			General temporal reference: e.g., "latest," "new," "recently published"	9	43	23	24	32	28

[a] Only one (specific) report contained partial data.

116 separate reports, 21 specific and 95 general. The analysis focused primarily on the information provided in each of the specific reports, but included the general reports as well (Table 2). It should be stressed that it is not at all clear what information *ought* to be provided in a general report of poll results. Therefore no evaluative conclusions should be drawn from Table 2 about the quality of general coverage. Instead, general reports are included to show the total range of information provided in all reports of poll results. It is also important to note that Table 2 shows the number of reports in which *any* information of a particular type was given. No attempt was made to determine whether the information provided was *adequate* by some standard. For example, 17 reports included information about "Timing," the date(s) on which interviewing was conducted. But in a majority of cases, the actual information was that a poll had been conducted "last week."

Table 2 shows that very little information was provided in reports of poll results, even if general reports are included. Judged by APC guidelines, specific reports were simply inadequate. Of 10 required items of information six were never mentioned and two were mentioned only once; no item was included in a majority of reports. At a more specific level of analysis, two findings are of particular interest. The first relates to the "Timing" category. As noted above, the entire official campaign lasted only 19 days, but there was an average time lag of 8 days between interviewing and first publication for the polls reported on television. Even the latest polls on television were conducted a full week before the date of the election. In these circumstances, information about timing is crucial for a correct interpretation of poll results, but this information was provided in only 9 (43 percent) of the specific reports. Instead, poll results were described in terms of their release date (57 percent) or an ambiguous temporal reference such as "latest" or "new" (43 percent). Both of these, of course, increase the apparent timeliness and newsworthiness of the results; each occurred in 28 percent of all reports. The second finding concerns the way in which poll data were reported. The 92 election items included only 21 specific reports of poll results. As shown in Table 2, only 10 of these included complete data. This is especially noteworthy in that complete data could be provided by reporting a single two-party preferred figure. Further, poll data were identified as first preference or two-party preferred in only six reports.

At least two major objections could be raised against this analysis. First, it can be argued that the various standards were devised with the print media in mind and that it is inappropriate to apply them to television with its special constraints and its emphasis on "secondary" reporting of results. Certainly, some television journalists (e.g., McKee, 1982) have devised and applied their own standards for acceptable coverage. This issue is examined in more detail elsewhere.[10] Second, it can be argued that the only reason for including methodological information is to assist the viewer in drawing accurate conclusions from the poll data. If television journalists provide valid interpretations of the results, then this information is unnecessary. This issue is examined next.

ERRORS IN COVERAGE

The fourth goal of the study was to assess the validity of television poll coverage. This presents two major analytical problems. First, the number of possible errors of fact and inference is indefinitely large. Put somewhat differently, it is impossible to specify in advance all of the ways in which something can be *mis*interpreted. Thus no truly comprehensive analysis of reporting accuracy is possible. The only alternative is to specify a number of possible reasoning errors and then examine instances of reporting to see if those errors occur. The second problem is the difficulty of specifying exactly what constitutes an "error" of reasoning. It was noted above, for example, that journalists in this study often drew future implications from poll results and attributed effects to polls without providing corroborating evidence. The propriety of these practices can be questioned, but even in specific instances it is seldom possible to demonstrate conclusively that they are erroneous. Recognized standards of inductive reasoning and social causation are simply not adequate for the purpose.

These considerations determined the analytical procedures used in this phase of the study. First, questionable poll reporting practices suggested by an analytical model of the poll inference process, prior research, and past experiece were listed and clarified. Nine were then selected for use in the analysis. Each of them clearly embodies a recognized error of fact or inference and each is marked by characteristic semantic or surface linguistic features. These possible errors are briefly summarized in Table 3. Finally, all poll-relevant material in the 92 poll items was examined and instances of the various errors were recorded. Each instance was further categorized as occurring in a direct statement by a journalist, politician, or other public figure (e.g., academic, businessman) or in a journalist's paraphrase of a statement by a named politician.[11] The results of this analysis are shown in Table 3.

A total of 203 errors of fact or inference were found in the 92 poll items. Given that reporting accuracy is a primary value in western journalism,

---

[10] Smith (1984) argues that abbreviated standards may be appropriate for secondary coverage, but the information provided must include population, interview date, question wording, and reference to the location of primary coverage.

[11] All errors in the video were coded under the "Journalist" heading.

Table 3. Errors in Coverage of Opinion Polls, 92 Items

	Error Source				
Error	*Journalist*	*Journalist Paraphrasing Politician*	*Politician*	*Other*	*Total*
Factual inaccuracy: clear misreporting of a factual detail, e.g., reporting incorrect poll data.	4	0	1	0	5
Overgeneralization: standard inductive fallacy, e.g., invalid generalization across polls, or generalization to a population different from that sampled.	15	0	0	0	15
Modal force: any overstatement of the strength of a claim, especially drawing *certain* conclusions from probabilistic poll data.	39	8	14	2	63
Conclusive swing: a species of modal force error, claim that some consequence would certainly (rather than probably or possibly) follow from a swing of some magnitude.	4	0	0	0	4
Temporal transposition: Reporting poll results in present (or future) terms, e.g., "The poll shows Labor is ahead," or "Today's poll shows Labor ahead."	50	0	1	0	51
Hypothesis contrary to fact: fallacy of drawing a conclusion from a contrafactual situation, e.g., "If the election had been held last week the ALP would have won."	8	0	2	0	10
Ambiguous comparison: reporting results in reference to an unspecified baseline, e.g., "The results show a 4% swing to Labor." Erroneous because unverifiable and *essentially* ambiguous.	29	5	0	0	34
Evaluative description: reporting poll results solely in evaluative (rather than factual) terms, e.g., "The latest poll is another good result for Labor."	13	1	2	0	16
Invalid interpretation: Selection or manipulation of results in an attempt to contradict a valid intepretation, e.g., averaging sequential poll results in order to contradict their sequential implications	5	0	0	0	5
Total, all errors	167	14	20	2	203

and that the analysis focused on only nine of the hundreds of possible errors, this figure does seem remarkably high. A large majority (82 percent) of the errors were committed by journalists. This finding is certainly compatible with Noelle-Neumann's (1980) claim that journalists often display an inadequate understanding of statistical data, but the finding may be somewhat misleading. Although it is true that journalists committed the great majority of errors, it is also true that journalists filled a great majority of the air time. When this factor is taken into account, the error rate for politicians appears to be at least as high as that for journalists. It should also be stressed that no attempt was made to trace the errors to their ultimate sources. In at least some instances, especially errors of hypothesis contrary to fact, it seems likely that journalists were merely repeating errors found in the news releases of polling organizations.

BALANCE OF COVERAGE

The final goal of the study was to determine the degree of balance in coverage of the strengths and weaknesses of the polls reported. Wheeler has claimed (1976:xvi–xvii) that "the reader of opinion polls is given no hint that they may not be trustworthy. The press in particular is oblivious to the pitfalls of polling." This is substantiated by Paletz et al., who note (1980:506) a "natural inclination of journalists to support rather than undermine the polls they report." Pressures to present the poll results as accurate were especially strong in 1980: Labor's unexpected lead in the polls allowed journalists to transform the campaign from a dull and predictable government walkover into an exciting horserace.

But there were several important reasons for interpreting the poll results cautiously. Aside from the unexpectedness of the findings and the problem of time lag noted above, the polls had consistently underestimated the coalition's share of the vote in elections held in the 1970s. This occurred in 12 of the 15 final polls during that period and was particularly striking in the 1977 election, the last before 1980. In that campaign, three of the four final polls "predicted" a narrow coalition victory with 51 percent of the two-party preferred vote; in the event, the coalition was returned to power with 54.6 percent of the vote and the second largest parliamentary majority in Australian electoral history. Based on these and other considerations, warnings against facile interpretation of the polls abounded in the print media in 1980 (e.g., Maley, 1980:308–9).

For this phase of the study, all relevant material in the 92 poll items was examined and comments supporting either rejection or acceptance of the poll results were noted. These comments were then categorized on the basis of the type of reason offered for rejecting or accepting the results.

Table 4. Reasons Given for Rejection or Acceptance of Poll Results, 92 Items

Reason	*Source* *Journalist*	*Journalist Paraphrasing Politician*	*Politician*	*Total*
A. Rejection				
1. Qualifier: any general reservation about poll accuracy, e.g., "If the polls are right . . . ."	1	2	4	7
2. Irrelevance: general statement that claims or implies poll results have no bearing on the election, e.g., "There's only one poll that counts"	1	2	5	8
3. General rejection: explicit rejection of polls or poll results without providing grounds for the rejection, e.g., "The polls are always misleading"	0	2	7	9
4. Ambiguity: claim that poll results have no clear implications or conflicting implications	1	0	0	1
5. Misinterpretation: claim that true results or interpretation differ from those given	0	0	2	2
6. Past Inaccuracy: claim that polls have been erroneous in past elections, Australia or elsewhere	0	0	0	0
7. Contradiction by other indicators: claim that other indicators (e.g., contributions, rally turnouts, mail, legislative performance) contradict poll results	2	1	1	4
8. Bias: claim that a polling organization is politically biased, therefore its results are biased	0	0	0	0
9. Dated: claim that poll results are no longer valid because of time elapsed since interviewing or because interviewing occurred before some crucial event	0	2	2	4
10. Methodological flaws: claim that poll results are invalid because of a methodological flaw such as sampling, question wording, distribution assumptions	0	0	0	0
11. Undecided voters: any specific mention of undecided voters as a possible source of invalidity	0	0	0	0
12. Past swings: any mention of late swings in previous elections as a possible source of invalidity	1	0	2	3
13. Possibility of swing: specific claim that a late swing could change results in this election	2	2	5	9

(continued)

Table 4. (Continued)

	Source			
Reason	*Journalist*	*Journalist Paraphrasing Politician*	*Politician*	*Total*
14. Assertion of swing: claim that election results will be different from poll results	2	1	6	9
Total, all rejection	10	12	34	56
B. Acceptance				
1. General Acceptance: explicit acceptance of polls or poll results without providing grounds for the acceptance, e.g., "Polls are the best guides we have"	0	0	0	0
2. Trend of results: claim that a poll result continues a trend of similar results and is therefore accurate	3	2	3	8
3. Convergence of results: claim that results of different polls conducted at the same point in time are similar and therefore accurate	7	0	0	7
4. Repetitive stress: a passing reference that links present poll results with others, e.g., "Yet another poll shows . . ."	26	2	0	28
5. Past Accuracy: claim that polls have been accurate in past elections, Australia or elsewhere	2	0	0	2
6. Confirmation by other indicators: claim that other indicators (e.g., contributions, rally turnouts, mail, legislative performance) confirm poll results	0	0	5	5
7. Prestige: claim that poll results are valid because of the good reputation of the polling organization responsible for them	0	0	0	0
8. Recency: claim that poll results are accurate because recent	2	0	1	3
Total, all acceptance	40	4	9	53

The typology used for this purpose was constructed by first listing all the grounds on which poll results might be criticized. These are briefly described in Table 4A. These criticisms fall into five groups: unsubstantiated rejection (items 1–3), problems of interpretation (4–5), invalidating factors external to the polls (6-7), invalidating factors internal to the polls (8–11), and possible changes in the electorate (12–14). The typology was completed by converting as many as possible of the grounds for rejection into corresponding grounds for acceptance. These are described in Table 4B.[12] Finally, all comments were further categorized on the basis of source.[13]

On the most general level of analysis, the results shown in Table 4 appear to contradict Wheeler's claim: rejection and acceptance comments were almost perfectly balanced, 56 to 53. A closer inspection, however, reveals an interesting pattern. Of the 56 rejection comments, 44 (79 percent) were either made by politicians or attributed to them. An examination of the transcripts reveals that many of these comments were presented as the predictable political response to unfavorable polls, and some were explicitly rebutted by journalists. In contrast, 40 (75 percent) of the 53 acceptance comments were made by journalists. None of these were rebutted. It should be noted, however, that a majority of the acceptance comments took the form of the relatively weak repetitive stress. Bearing this in mind, the results still provide strong support for the claim of a "natural inclination of journalists to support rather than undermine the polls they report." It is also useful to note the types of reasons given for accepting or rejecting the poll results. Despite the experience of 1977 and the existence of substantial differences among the polls in 1980, a large majority of acceptance comments stressed the similarity of the poll findings (Table 4B, items 2–4). Nearly all the rejection statements, on the other hand, took the form of unsubstantiated rejections (4A, items 1–3) or discussions of the possibility of a late swing (items 12–14). Specific criticisms of the results (items 4–11) were seldom mentioned. In particular, only four comments, none by journalists, pointed to the crucial problem of dated results; past inaccuracies, methodological flaws, and undecided voters were mentioned by no one.

Discussion

The results of this study confirm earlier American findings that poll coverage is extensive, superficial, and inaccurate. The problem, then, is to account both for this pattern and for its cross-cultural occurrence. One possible explanation is based on the claim that television news is the product of purposive human action (Tuchman, 1978). All human action, in turn, is shaped by the actor's perceptions of the constraints and opportunities in a given situation (Smith, 1980). More specifically, we would expect television journalists to recognize at least four major constraints. Given strict time limitations, the primarily oral and sequential nature of television news, and the need to inform a heterogeneous audience, coverage must be brief and therefore *simple*. Further, prevailing journalistic standards require *timely*, *interesting*, and *accurate* coverage. If these constraints are assumed,the results of this study form a coherent pattern.

---

[12] Only positive reasons for accepting poll results (e.g., the poll is accurate because it is recent) were listed. Possible rebuttals of reasons for rejection (e.g.,the poll is accurate because it is not methodologically flawed) were not included. Hence the list of reasons for acceptance is smaller than the list of reasons for rejection. In the event, no rebuttals were noted.

[13] The "Other" category was empty and so is omitted from Table 4.

The extensive reporting of opinion polls is clearly explicable as an attempt to enliven mandatory coverage of a dull and predictable election campaign. But a major problem arises here. Opinion polls are inherently complex, somewhat dated, and tentative in their conclusions; television news should be simple, timely, and interesting. How can this obvious incompatibility be resolved? One way is to transform the poll findings so that they match the requirements of television news. This requires three operations: *simplification*, *temporal transposition*, and *dramatization*. Each of these is evident in the results.

Simplification was achieved by omitting methodological details, reporting results in general terms, and committing the errors of overgeneralization, ambiguous comparison, and evaluative description. Temporal transposition was accomplished by stressing the recency of the polls, discussing their meaning in terms of present or future implications, and committing the error of temporal transposition. Dramatization took two forms. First, through the errors of modal force and hypothesis contrary to fact, the tentative and qualified claims of the pollsters were transformed into conclusive statements of fact. These "facts" were then supported by dispassionate journalists, rejected only by self-serving politicians. Second, the polls were depicted as powerful causal agents. And in so far as politicians were forced by journalists to react to the polls,they became such agents.

The result of these operations, of course, is that poll reporting was inaccurate. This may have been due to conscious choice, but it seems more reasonable to conclude with Noelle-Neumann (1980) that the journalists simply lacked sophisticated knowledge of polling techniques and interpretive procedures. Guided by habitual criteria of newsworthiness, they would have been unaware of their distortions.

Because these constraints are shared by most Western journalists, we would expect the pattern of coverage discovered here to be widespread. Thus, it becomes necessary to determine what damaging effects it might have. One concern is its impact on the electoral process, especially the decisions of "tactical" voters (Worcester, 1980). A second concern is the effect on polling itself, and here the Australian experience is instructive.

In Victorian television coverage of the 1980 election, polls were stripped of their protective qualifiers and reservations and presented as a dominant focus of interest. Unfortunately, three of the four failed to "pick the winner." But with one exception (Saulwick's *Herald/Age* survey), there is no reason to think the polls were seriously flawed. Indeed, three improved their margin of error over the 1977 election. What seems to have happened is that a small shift back to the Government occurred very late in the campaign, after the final interviews had been conducted. This tipped the scales in a close election.

But these technicalities seldom emerged in postelection coverage. The polls were vehemently denounced as inaccurate and misleading by politicians and journalists alike. Perhaps more disturbing, new elections were held in March 1983. Tapes are available for one Victorian station (Channel 2) and they reveal dramatic changes in coverage. In 1980, Channel 2 broadcast 21 poll items: 2 headlines, 2 mentions, 10 themes, and 7 focuses; in 1983, only 4 were broadcast, all mentions. Somewhat more coverage was provided on the evening public affairs program "Nationwide," but it stressed the limitations of polls.[14] It seems reasonable to conclude that public confidence was diminished by these events. And as argued above, that confidence is a necessary condition for successful polling.

If this reporting pattern is widespread and damaging, efforts to change it are urgently required. The first step is to determine what constitutes "accurate" poll reporting. This requires development of guidelines that specify the minimum information needed in "secondary" reports of poll results (Ladd, 1980; Smith, 1984) and that clarify the range of valid inferences that can be drawn from poll data. The next step is to train journalists in their use. It must be stressed, however, that the conflict between the nature of opinion polls and the constraints of television journalism is real and fundamental. Thus it may be the case that television news *cannot* provide accurate coverage of opinion polls and other complex and subtle events. This, in turn raises fundamental questions about the adequacy of television as a primary source of news.

References

Barter, John
1982 "Newspaper coverage of the polls." Pp. 126–31 in Robert Worcester and Martin Harrop (eds.), Political Communications. London: George Allen and Unwin.
Broh, C. Anthony
1980 "Horse-race journalism: reporting the polls in the 1976 presidential election." Public Opinion Quarterly 44:514–29.
Cantril, Albert H. (ed.)
1980 Polling on the Issues. Cabin John, MD: Seven Locks Press.
Crespi, Irving
1980 "Polls as journalism." Public Opinion Quarterly 44:462–76.
Crewe, Ivor
1981 "As the world turns out." Public Opinion 4:52–53.

---

[14] Viewers were told, for example, that "during the last election most of the opinion polls were way out" and that "this poll is not the definitive measure of the state of public opinion at the moment. The poll result is a guide only, a little empirical evidence to go along with all the other information that you and I use to guess what might be the result on March 5."

1982 "Improving but could do better: the media and the polls in the 1979 general election." Pp. 115–25 in Robert Worcester and Martin Harrop (eds.), Political Communications. London: George Allen and Unwin.

FACTS (Federation of Australian Commercial Television Stations)

1980 FACTS Report 1980. Sydney: FACTS.

Germond, Jack W.

1980 "The impact of polling on journalism." Pp. 20–27 in Albert H. Cantril (ed.), Polling on the Issues. Cabin John, MD: Seven Locks Press.

Krippendorff, Klaus

1980 Content Analysis. Beverly Hills, CA: Sage.

Ladd, Everett C.

1980 "Polling and the press: the clash of institutional imperatives." Public Opinion Quarterly 44:574–84.

Ladd, Everett C., and G. Donald Ferree

1981 "Were the pollsters really wrong." Public Opinion 3:13–20.

Mackerras, Malcolm

1980 Elections 1980. Sydney: Angus and Robertson.

Maley, Bill

1980 "The opinion polls." Pp. 301–10 in Malcolm Mackerras, Elections 1980. Sydney: Angus and Robertson.

McKee, Paul

1982 "ITN's use of opinion polls." Pp. 132–40 in Robert Worcester and Martin Harrop (eds.), Political Communications. London: George Allen and Uwin.

Miller, M. Mark, and Robert Hurd

1982 "Conformity to AAPOR standards in newspaper reporting of public opinion polls." Public Opinion Quarterly 46:243–49.

Noelle-Neumann, Elisabeth

1980 "The public opinion research correspondent." Public Opinion Quarterly 44:585–97.

Paletz, David L., et al.

1980 "Polls in the media: content, credibility and consequences." Public Opinion Quarterly 44:495–513.

Pryor, Geoff

1980 "Press buries the issues." New Journalist 36:14–17.

Smith, Ted J.

1980 "On the nature and sources of practical necessity." Philosophy of the Social Sciences 10:379–96.

1981a "Television news and the 1980 Australian federal election campaign." Presented to the Second National Conference of the Australian Communication Association, Sydney, May.

1981b "Opinion polls, political demonstrations, and television coverage of the 1980 federal election." Regional Journal of Social Issues 9:13–28.

1984 "Minimum standards of disclosure for television coverage of opinion polls." Presented to the 39th Annual AAPOR Conference, Delavan, WI, May.

Snider, G. A.

1976 "An evaluation of various measures of swing voting in Australia, 1969–1975." Politics 11:185–89.

Thompson, Elaine

1981 "Public opinion polls, statistics and bias." Current Affairs Bulletin 53:28–31.

Tuchman, Gaye

1978 Making News: A Study in Construction of Reality. New York: Free Press.

Wheeler, Michael

1976 Lies, Damn Lies, and Statistics. New York: Liveright.

1980 "Reining in horserace journalism." Public Opinion 3:41–45.

Worcester, Robert M.

1980 "Pollsters, the press, and political polling in Britain." Public Opinion Quarterly 44:548–66.

The Failure of Liberal/Conservative Ideology as a Cognitive Structure

NORMAN R. LUTTBEG AND MICHAEL M. GANT

EARLY voting research largely dismissed the public's ability to use the ideological labels of liberal and conservative in structuring their political thoughts, as well as in evaluating issues and candidates. Herbert McClosky (1958:44) states, "the correlation(s) between them [attitudes on policy issues] tend . . . to be fairly low, suggesting that for the present, at least, many Americans divide in their party preferences, their support of candidates, their economic views, their stands on public issues, or their political self-identification without reference to their beliefs in liberalism or conservatism." The authors of *The American Voter* note (Campbell, et al., 1960: 543), "Our failure to locate more than a trace of 'ideological' thinking in the protocols of our surveys emphasizes the general impoverishment of political thought in a large proportion of the electorate." Most of the research since *The American Voter* has focused on two questions: Has the ability of the American citizenry to think in ideological terms improved since the quiet, issueless 1950s? And, can the

Abstract While the American electorate has been found less than fully capable of choosing among presidential candidates on the basis of their positions on the issues, there has been optimism that were that electorate encouraged to conceptualize politics on the basis of liberal and conservative ideological positions, the quality of vote choice would improve. Using ANES 1980's questions to assess the electorate's ability to lend correct definitions to the terms *liberal* and *conservative*, we find little basis for optimism that an ideologically sophisticated electorate would better select candidates reflecting the electorate's positions on the issues.

Norman R. Luttbeg is a Professor and Michael M. Gant is an Associate Professor in the Department of Political Science, Texas A&M University. The data utilized in this paper were made available by the Inter-University Consortium for Political and Social Research. Neither the collectors of the original data nor the Consortium bear any responsibility for the analysis or interpretations presented here.

Public Opinion Quarterly Vol. 49:80–93 © by the Trustees of Columbia University
Published by Elsevier Science Publishing Co., Inc. 0033-362X/85/0049-80/$2.50

ability of the American public to think abstractly be assessed validly with more sensitive and rigorous methodologies?

Many techniques have been used to assess public facility with ideological thought. Respondent articulation of the words *liberal* or *conservative* in response to the "likes/dislikes" questions used since 1952 in the Survey Research Center (now called Center for Political Studies) American National Election Studies, while the most commonly used measure (Campbell, et al., 1960; Converse, 1964; Repass, 1971; Hagner and Pierce, 1982; Hill and Luttbeg, 1981), does not directly assess the ability of the public to use the words *liberal* and *conservative* in a meaningful way. Another method of measuring ideological thinking involves the assessment of "correct" interrelationships between issue positions and other attitudes (McClosky, 1958; Converse, 1964, 1970), but this method is substantially affected by question wording (Bishop, et al., 1978; Sullivan, et al., 1978). People's willingness to identify themselves as liberal or conservative has also been used, despite doubts concerning whether such self-identification facilitates issue-based voting judgments (Levitin and Miller, 1979; Conover and Feldman, 1981).

Our purpose in this paper is twofold. First, we wish to assess a more direct measure of the public's familiarity with the terms *liberal* and *conservative*, a measure which appeared in the 1980 CPS American National Election Studies.[1] Second, we wish to ascertain whether or not the ability to define these terms, and to define them correctly, aids the citizen in engaging in issue voting. By *issue voting*, we mean voting for the candidate closest to the citizen in the issue space. This simple spatial proximity approach derives from rational choice theories of electoral choice (cf. Downs, 1957; Stokes, 1966).

The measure used here to assess what people mean by the terms *liberal* and *conservative* is the following question from the 1980 CPS survey:

> People have different things in mind when they say that someone's political views are liberal or conservative. We'd like to know more about this. Let's start with liberal. What sort of thing do you have in mind when you say that someone's political views are liberal (conservative)?

An earlier filter question,

> We hear a lot of talk these days about liberals and conservatives. Here is a seven-point scale on which the political views that people might hold are arranged from extremely liberal to extremely conservative. Where would you place yourself on this scale,or haven't you thought much about it?

---

[1] Somewhat similar questions were asked in the United States in 1974 in a comparative study reported by Hans Klingemann (1979). For comparative reasons, "left" and "right" were substituted for liberal and conservative and meaning assessed differently. Given substantially different operationalizations, there is little in this study which can be cited here to note change over time or comparisons with other democracies.

had found 441 respondents of the 1400 taking the postelection wave survey saying either that they had not thought much about the words, or did not know what they meant. The most common first mention definitions of both *liberal* and *conservative* from the 1980 survey are shown in Table 1. From the multiple responses to this question (up to three definitions were solicited), we constructed a measure of accuracy in using these terms.

Compared to the findings of the 1950s, Table 1 portrays a fairly sophisticated electorate. Substantial proportions of the characteristics attributed to liberalism and conservatism have to do with the abstract notions of change, economic perspectives, and approaches to problem solving. These data do not lead to the conclusion that the public is now capable of ideological thinking. They do, however, strongly suggest that substantial numbers of people can articulate characteristics of liberalism and conservatism which are consistent with the terms of public debate.

And yet, Table 1 cannot inform us as to the accuracy of these definitions, the correlates of accurate use of these terms, or the utility of ideological thinking for self-interested voting. We turn, then, to the following questions: (1) What are the determinants of the ability to use the terms *liberal* and *conservative* correctly? (2) Do those who can and cannot use these terms accurately differ in their ability to discover, and vote for, the candidate whose policy positions are closer to their own?

Table 1. Distribution of First Definitions of *Liberal* and *Conservative* (N = 1400)

Definitions	%
Liberal	
Acceptance of change and new ideas	10.6
Spend more freely; favor government spending	7.4
Social welfare; "give-away programs"	4.9
Socialistic, for welfare state: less to private enterprise	4.7
Quick (rash) response to problems	2.4
Other responses	30.0
Haven't thought about	31.5
No answer or don't know	8.5
Conservative	
Resistance to change and new ideas	10.3
For free entrprise, capitalism	4.6
Slow (cautious)	4.3
Spend little; less wasteful	3.7
For states' rights	2.4
Other responses	31.5
Haven't thought about it	31.5
No answer or don't know	7.4

Accuracy of Liberal/Conservative Definitions

To construct our measure of accuracy of liberal/conservative definitions, we began by recoding the responses to these questions, collapsing all responses to each variable into five possible categories. Our recoding exercise was based on three premises. First, we assumed that it is possible to distinguish between correct and incorrect definitions of the words *liberal* and *conservative*. Moreover, we assumed that attributes which are correct descriptions of liberalism are incorrect descriptions of conservatism, and vice versa. For example, responses which indicated support for greater spending at the national level, exclusive of national defense spending, were coded as correct definitions of liberal and incorrect definitions of conservative.

Second, we assumed that correct definitions of *liberal* and *conservative* should be consistent with enduring understandings of these terms in the context of American politics. Until the election of Ronald Reagan and the advent of "new federalism," "deregulation," and other programs of conservative change, an individual would be correct in perceiving acceptance of change as characteristic of liberals. We therefore categorized as "incorrect" the response "acceptance of change" as an attribute of conservatism, and the response "resistance to change" as an attribute of liberalism.

Finally, we assumed that the ability to define these terms should be helpful in leading an individual in voting for that candidate whose policy positions are closer to his/her own positions (cf. Downs, 1957: 4–14). If a response was judged to be noninstrumental to selecting a government that would follow policies favored by the respondent, e.g., the responses "independence of thought" or "concern for public morality," that response was categorized as "extraneous."[2]

Beyond the categories "correct," "incorrect," and "extraneous," three other categories were created. The first is the "missing data" category, which consists of those respondents who did not participate in the postelection wave of the survey. The category "no content" consists of those respondents who could not ascribe a definition to the words *liberal* or *conservative*. Finally, several responses in the protocol published in the codebook for the 1980 CPS American National Election Study were so vague as to defy judgment as to the correctness of the definition. These responses were categorized as "insufficient." A complete listing of the codes for these variables subsumed under each of these categories is given in the Appendix.

---

[2] Our categorizing "concern for public morality" as an extraneous consideration should not suggest that no public policies are concerned with moral issues. For example, abortion policies and school prayer occupy important places on the current political agenda. But the coding "public morality" leaves secondary analysts uncertain as to what the respondent said. Given this ambiguity, and the impossibility of ascertaining either the correctness or relevance of this category of responses, we have labeled them "extraneous," which implies neither correctness nor incorrectness. Moreover, it is impossible to construct an argument to demonstrate that either liberals or conservatives are more concerned with public morality.

We next created two variables, one from the liberal responses, and one from the conservative responses. The categories for these variables are as follows:

No Wrong: At least one of the three responses is correct, and no response is incorrect.

No Right: At least one of the responses is incorrect, and no response is correct.

No Content: No response on all three definition variables.

Other: All other mixes of responses not detailed above.

Missing: No postelection interview.

The final step in the operationalization process involved the creation of a variable which is a combination of the previous two measures. In addition to those who did not participate in the postelection survey, those whose only responses were "other" on these measures were classified as "missing." The other categories are "both correct," which consists of no incorrect definitions of *liberal* or *conservative*; "one or both incorrect," which consists of no correct definition of *liberal* or *conservative*; and "no content," which consists of individuals who could offer definitions for neither *liberal* nor *conservative*. The distributions for all three measures are presented in Table 2.

Sizable proportions of respondents to the 1980 CPS survey could articulate no attributes of either liberal or conservative. Those articulating a response, however, were usually correct in their definitions of liberal and conservative (80 percent for both concepts). In general, about two out of five respondents could not venture a description of these words. Of those who did, however, most were correct.

Correlates of Ability to Define Liberal and Conservative

We have demonstrated that large numbers of respondents to the 1980 CPS survey were unable to define the terms *liberal* and *conservative*. Yet, of those who did attempt to define these terms, most were correct in their assessments. Before we explore the implications of this knowledge, or lack of knowledge, for issue voting, we wish to assess some correlates of the ability to define, and define correctly, the attributes of liberalism and conservatism.

As might be expected, the ability to define the terms *liberal* and *conservative* is a function of education (cf. Table 3). The relationship is

Table 2. Distributions of Liberal Accuracy, Conservative Accuracy, and Ideological Accuracy

	f	%
Liberal accuracy		
No incorrect	670	48
No correct	22	2
Other	148	11
No content	560	40
Total	1400	
Missing	214	
Conservative accuracy		
No incorrect	683	49
No correct	16	1
Other	156	11
No content	545	39
Total	1400	
Missing	214	
Ideological accuracy		
Both correct	569	42
One or both incorrect	248	18
No content	531	39
Total	1348	
Missing	266	

Table 3. Accuracy of Ideological Definitions by Level of Education

	Education				
Accuracy of Definitions	*Less than HS Diploma*	*HS Diploma*	*Some College*	*College Degree or More*	*Row Totals*
Both correct	19	38	57	68	42
One or both incorrect	13	18	21	25[a]	18
No content	68	44	22	7	40
(N)	(354)	(473)	(287)	(231)	(1345)

NOTE: Entries are column percentages.
$\chi^2 = 264.2, p < .0001$
Gamma = −.53
Gamma for first two rows, all columns = −.15
[a] χ^2 for first two rows, all columns = 9.4, $p < .05$

moderately strong, in the correct direction, and statistically significant. It is clear, however, that most of the observed relationship is due to the distinction in willingness to hazard a definition; there is only a weak relationship between level of education and the ability to define these terms *correctly*. While college graduates are much more likely to articulate a definition of liberalism and conservatism (93 percent), they are apt to be wrong in these assessments (25 percent) as well. When those classified as "no content" are excluded from this table, the relationship is weakened considerably (gamma = −.15), and barely achieves statistical significance $\chi^2 = 9.4; p < .05$).

A similar pattern was found when we compared males with females, and whites with nonwhites, in terms of the ability to define *liberal* and *conservative*, and to do so correctly. Men were better able to define these terms than women. Similarly, white respondents were more likely to offer definitions than nonwhites. But as with education, when those classified as "no content" are excluded from the analysis—i.e., when respondents are compared only in terms of definitional accuracy—the relationships largely disappear. In the comparison between males and females, the relationship is not significant. In the comparison between whites and nonwhites the relationship actually reverses direction, but again the results are not statistically significant. Finally, when we examined the impact of strength of partisanship on definitional accuracy, we found no relationship. Thus, we have found some correlates of venturing a definition, but few of being *correct* in that definition.

Definitional Accuracy and Issue Voting

We turn now to the central question: Does the ability to correctly define the terms *liberal* and *conservative* aid citizens in identifying and voting for the candidate closer to them on the issues of the campaign? We earlier defined this phenomenon as "issue voting." This definition follows from the assumptions that the purpose of an election is to select a government, that policy making is the purpose of that government, and that rational citizens will attempt to discover those candidates more likely to pursue policies favored by them, and will vote for those candidates.

As we might expect, respondents who could not offer a definition of *liberal* or *conservative* voted in substantially lower numbers (with a turnout rate of 43 percent) than those who could define these terms (with a turnout rate of 65 percent). But accuracy of these definitions was inconsequential to turnout, as both those who were correct and those who were incorrect in their definitions voted at the same 65 percent rate. With these results in mind, let us more fully explore the relationship between issue voting and accuracy of ideological definitions.

To measure issue voting we constructed measures of candidate preference based on each of the "seven-point" issue scales from the 1980 CPS National Election Study.[3] For each of the seven scales the following calculation was performed:

---

[3] The seven-point issue scales used in this analysis are: Government Spending and Services; Defense Spending; Guaranteed Jobs and Standard of Living; Aid to Minorities; Relations with Russia; Women's Rights; and Inflation/Unemployment. Information re-

$$(| X_i - \bar{\theta}_{Ci} | - | X_i - \bar{\theta}_{Ri} |)$$

where X_i is the respondent's position on an issue scale, and $\bar{\theta}_{Ci}$ and $\bar{\theta}_{Ri}$ are the *mean perceived* positions of Carter and Reagan on that scale. We use the mean to represent the candidates' "true" positions on these scales in order to avoid the contaminating effects of such selective mechanisms as projection and persuasion (cf., Zajonc, 1971; Brody and Page, 1972). In the absence of detailed information regarding the "true" positions of candidates, the mean is often used as a substitute (cf. Bruner, 1978; Marcus and Converse, 1979).

This measure of spatial proximity was calculated for all respondents who placed themselves and both candidates on a scale; if these conditions did not hold, a score of zero was assigned to the respondent. Vote choices were then predicted as a function of spatial proximity on each scale, and on all seven scales, using probit analysis.[4] The reported and predicted vote for each respondent who voted, and voted for either Carter or Reagan ($N = 783$; validated voters only), were then compared. Those whose reported and predicted votes matched are called *issue voters*. The proportions of the validated voters who are issue voters ranged from 50 percent for Women's Rights to 66 percent for Defense Spending and Guaranteed Jobs.

Fifty-one percent of the 783 validated voters cast votes for Reagan. With the exception of the Women's Rights scale, each issue scale predicts substantially better than this 51 percent expectation. Not surprisingly, the most accurate predictions are made using all seven issue scales; 75 percent of the respondents are classified as issue voters based on all scales. The following analysis is based largely on the "total" issue voter

---

garding these variables and other aspects of the 1980 American National Election Study can be gained from the codebook for the survey, published by the Inter-University Consortium for Political and Social Research, Ann Arbor, Michigan; Warren E. Miller, principal investigator.

[4] Probit analysis is a technique, similar in intent and interpretation to regression, which is designed to allow estimation of categorical dependent variables. It is therefore a more appropriate technique for estimating vote choice, a dichotomous nominal level variable.

The probit program which we utilized provides, in addition to estimates of coefficients for each of the independent variables, a "prediction" for each unit of analysis on the dependent variable. That is, on the basis of the probit function estimated for the entire sample, and the values on the independent variable for each respondent, a Carter vote or Reagan vote was "predicted" for each individual in the sample. These are the predicted votes referred to in the text. Our measure of issue voting was constructed using the predicted votes based on all seven issue scales. The probit function resulting from this analysis yielded correct predictions for 75 percent of the cases, with a correlation of predicted and actual values of rho = .49.

For a technical and nontechnical introduction to probit analysis see, respectively, McKelvey and Zavoina (1975) and Aldrich and Cnudde (1975).

measure, although we will have occasion to analyze the specific issue scale measures.

Table 4 presents the relationship between issue voting based on all seven issue scales and accuracy of liberal/conservative definitions. The absence of a relationship between issue voting and definitional accuracy is striking. While 77 percent of those who can correctly define both *liberal* and *conservative* are issue voters, 76 percent of those who define one or both terms incorrectly cast their ballots for the spatially proximate candidate. Even more striking is our finding that fully 72 percent of those who cannot or will not define these terms vote for the candidate closer to them on all issues taken as a whole. Simply stated, being able to define, correctly or not, the terms liberal and conservative is of no aid in guiding citizens toward a vote for the spatially proximate candidate.

The measure of issue voting in Table 4 reflects choices based on all seven issue scales. It may be that this measure masks differences in levels of issue voting which exist for some of the individual issue scales. In particular, those issues whose options reflect "traditional" liberal/conservative policy disagreements (Government Spending; Government Guaranteed Jobs; Aid to Minorities; and Inflation/Unemployment) may serve to discriminate between those who correctly and incorrectly define *liberal* and *conservative* relative to issue voting. This notion is examined in Table 5 where the level of issue voting based on each issue scale is presented for each category of our definitional accuracy measure. Respondents had to have placed themselves and both candidates on a scale to be included in these calculations.

On only three of these scales are the observed differences in issue voting statistically significant. Two of the scales, Government Guaranteed Jobs and Aid to Minorities, reflect "traditional" liberal/conservative distinctions. The third, Relations with Russia, does not. For all three scales the statistically significant differences are due to the relatively low level of issue voting among those who could not venture a definition of

Table 4. Issue Voting Based on All Proximity Measures by Accuracy of Liberal/Conservative Definitions; Validated Voters Only

	Accuracy of Definitions			
Issue Voting	*Both Correct*	*One or Both Incorrect*	*No Content*	*Row Totals*
Yes	77	76	72	75
No	23	24*	28	25
(*N*)	(348)	(153)	(248)	(749)

χ^2 for entire table = 1.7, $p < .4$

* χ^2 excluding "no contents" = .21, $p < .5$

NOTE: Entries are column percentages.

Table 5. Proportions of Issue Voting Based on Each Spatial Proximity Measure, by Accuracy of Liberal/Conservative Definitions; Validated Voters Only[a]

	Proportion of Issue voting by Accuracy of Definitions		
Issue Scale	*Both Correct*	*One or Both Incorrect*	*No Content*
Defensive	67	65	60
(*N*)	(303)	(129)	(149)
Government Spending	74	74	73
(*N*)	(285)	(113)	(124)
Guaranteed Jobs	72	71	63*
(*N*)	(290)	(120)	(126)
Minorities	69	72	56*
(*N*)	(304)	(125)	(128)
Russia	63	67	52*
(*N*)	(289)	(134)	(129)
Women's Rights	54	54	51
(*N*)	(291)	(129)	(148)
Inflation	66	66	68
(*N*)	(200)	(85)	(83)

NOTE: Entries are proportions of each category classified as issue voters.

* χ^2 significant at $p \leq .05$

[a] Only those placing themselves and both candidates are included for each scale.

liberal and *conservative*. Variation in issue voting between those who define these terms correctly and incorrectly is virtually nonexistent.

Since definitional accuracy has few correlates and issue voting is not a function of definitional accuracy, we might suspect that persons preoccupied with a limited number of salient issues, prototypically conceived as "single issue" voters, might achieve issue voting despite an inability to define the terms *liberal* and *conservative* (Natchez and Bupp, 1970; Repass, 1971). The relationship between liberal/conservative definitional accuracy and the number of issue scales on which respondents placed themselves and both Carter and Reagan is a substantial one ($\chi^2 = 101, p < .0001$). Yet the small number of issues employed by those who cannot offer a definition of either *liberal* or *conservative* (the "no contents") largely explains this relationship. When these respondents are removed, the relationship weakens greatly, and is no longer statistically significant. There is no relationship between the number of issues considered by a voter and definitional accuracy, a result consistent with those previously reported.

In Table 6 we examine the impact of the number of issue scales considered on levels of issue voting. Two things are apparent from this table. First, among validated voters for Carter and Reagan, a sizable proportion (80 percent) were able to articulate a position for themselves and both candidates on four or more issues. Second, while issue voting

Table 6. Issue Voting by Number of Issue Scales Used by Respondents; Validated Voters Only

	Number of Issue Scales				
Issue Voting	*One*	*Two*	*Three*	*Four or More*	*Total*
Yes	67	63	72	76	74
No	33	37	28	24	26
(*N*)	(33)	(57)	(61)	(597)	(748)

NOTE: Entries are column percentages
$\chi^2 = 5.62$, df = 3, $p > .10$

decreases among those concerned with fewer issues, the relationship is not statistically significant. Examining the "few-" or "single-" issue voters, then, little affects the patterns noted previously.

Discussion and Conclusions

Exploring the capability of the American public to think in ideological terms has enjoyed a long tradition. Before we assess the implications of our findings for this body of work, it would be beneficial to step back for a moment and consider the question of why ideology has so consumed our interest as students of electoral decision making.

It is obvious that, at one level, the concern with ideology, or belief systems, flows from the recognition that human beings are limited information processors (cf. Miller, 1956; Macaluso, 1975; Inbar, 1979; Gant and Davis, 1984). If voters in electoral campaigns are to use available information effectively, some organizational structure would seem necessary. Such a structure would simultaneously aid the citizen in processing and storing information about candidates and issues.

Moreover, an ideology, defined as "a particularly elaborate, close-woven, and far-ranging structure of attitudes" (Campbell, et al., 1960: 192) would seem to serve these functions nicely. Of course, not just any ideology would seem desirable. Given that this organizing structure would be used for processing political information, it must have relevant political content. Specifically, given that political debate in this country (as in most Western democracies) is phrased in liberal/conservative, or left/right, terms (Klingemann, 1979), an ideology should be similarly constructed.

Our findings, at one level, simultaneously confirm and contradict these expectations. While we do find that large numbers of respondents cannot define the terms *liberal* and *conservative*, those who do venture a definition are mostly correct. Many, thus, seem prepared to use ideology to structure information in an electoral campaign.

On another level, however, our findings call into question the very notion that an ideology structured in liberal/conservative terms is necessary to linking public preferences to government action. The proportion of those who could not accurately define *liberal* and *conservative* and who voted for the spatially proximate candidate is virtually identical to that of those accurately defining these terms. This suggests, of course, that familiarity with liberalism and conservatism is not necessary in structuring public debate for voters.

This need not mean that an organizing structure for incoming information is unnecessary, only that such a structure need not be fashioned along liberal/conservative lines. It may be the case, as Kessel (1972) has noted, that information about candidate issue positions may be processed using a direct spatial proximity cue, without reference to specific left vs. right distinctions (see also Campbell, et al., 1960: 205). Or, ideologies may be present among the inarticulate mass public, but are unrecognizable as such when conventional, survey-based techniques are employed (cf. Lane, 1959, 1962; Brown, 1970).

Information about candidate issue positions may also be organized around specific issues (cf. Natchez and Bupp, 1970; Repass, 1971), rather than on a left/right continuum. Those citizens who care strongly about one or a few issues need only know the candidates' positions on these policies; ideology would be irrelevant. Some suggestive, although highly tentative, evidence emerged from our analysis regarding this point. If we isolate those individuals who considered fewer than four issues in arriving at a vote choice, who voted for either Carter or Reagan, and who were *incorrect* in their definition of *liberal* and/or *conservative* ($N = 20$), 80 percent were classified as issue voters. While the number of respondents is small and the conditions highly restrictive, this result further strengthens our proposition that liberal/conservative ideology may be irrelevant to voting for the candidate closest to the citizen in the policy space.

We do not wish to overstate our findings; generalizations simply cannot be built on the basis of one election study. But this paper does suggest, and suggests rather strongly, that much of the "traditional" research regarding the ideological capability of the American public is simultaneously misleading and unduly pessimistic.

Appendix

The following codes for the ideological definition questions were classified as correct definitions of liberalism, and incorrect definitions of conservatism: 10 20 21 40 50 60 61 80–83 90 210 220 270 280 284–286 330 331 340 350 400–402 404 405 512 600 601 610 620 630 631 640 642 650 670 671 672 681–684 686 850 870.

The following codes were classified as incorrect definitions of liberalism, and correct definitions of conservatism: 110 120 121 140 155 160 161 180–183 190 230 231 240 250 283

310 320 370 371 384–386 412 500–505 641 660 700 701 710 720 730 731 740 742 770–772 781 782 784 786 840.

The following codes were classified as extraneous definitions of liberalism and conservatism: 30 35 70 71 84–86 130 135 170 171 184–186.

The following codes were classified as insufficient definitions of liberalism and conservatism: 150 200 212 260 281 290 299 300 312 360 390 490 643 690 744 750 760 780 785 790 800 810 820 830 860 880 881 890 891 900–903.

References

Aldrich, John, and Charles F. Cnudde

1975 "Probing the bounds of conventional wisdom: a comparison of regression, probit and discriminant analysis." American Journal of Political Science 19:571–608.

Bishop, George F., Alfred J. Tuchfarber, and Robert W. Oldendick

1978 "Change in the structure of American political attitudes: the nagging question of question wording." American Journal of Political Science 22:250–69.

Brody, Richard A., and Benjamin I. Page

1972 "Comment: the assessment of policy voting." American Political Science Review 66:450–58.

Brown, Steven R.

1970 "Consistency and the persistency of ideology: some experimental results." Public Opinion Quarterly 36:60–68.

Bruner, Jere

1978 "Make your own Nixon: projection of candidate issue stand as a fly in the ointment of policy voting." Paper delivered at the annual meeting of the Midwest Political Science Association.

Campbell, Angus, Philip E. Converse, Warren E. Miller, and Donald E. Stokes

1960 The American Voter. New York: John Wiley and Sons.

Conover, Pamela J., and Stanley Feldman

1981 "The origins and meaning of liberal/conservative self-identifications." American Journal of Political Science 25:617–45.

Converse, Philip E.

1964 "The nature of belief systems in mass publics." In David Apter (ed.), Ideology and Discontent. New York: Free Press.

1970 "Attitudes and non-attitudes: continuation of a dialogue." In Edward R. Tufte (ed.), The Quantitative Analysis of Social Problems. Reading, Mass.: Addison-Wesley

Downs, Anthony

1957 An Economic Theory of Democracy. New York: Harper and Row

Gant, Michael M., and Dwight F. Davis

1984 "Mental economy and voter rationality: the informed citizen problem in voting research." Journal of Politics 46:132–53.

Hagner, Paul R., and John C. Pierce

1982 "Correlative characteristics of levels of conceptionalization." Journal of Politics, 44:779–807.

Hill, David B.,and Norman R. Luttbeg

1981 Trends in American Electoral Behavior. Itasca, Ill.: F. E. Peacock.

Inbar, Michael

1979 Routine Decision-Making: The Future of Bureaucracy. Beverly Hills: Sage.

Kessel, John H.

1972 "Comment: the issues in issue voting." American Political Science Review 66:459–65.

Klingemann, Hans D.

1979 "Measuring ideological conceptualizations." In Samuel H. Barnes and Max Kaase (eds.), Political Action: Mass Population in Five Western Democracies. Beverly Hills: Sage.

Lane, Robert E.
1959 Political Life. Glencoe, Ill: The Free Press.
1962 Political Ideology. New York: Macmillan
Levitin, Teresa E., and Warren E. Miller
1979 "Ideological interpretations of presidential elections." American Political Science Review 73:751–71.
Macaluso, Theordore F.
1975 "Candidate image: personal attribute and issue content." Paper delivered at the annual meeting of the Southern Political Science Association.
Marcus, Gregory B., and Philip E. Converse
1979 "A dynamic simultaneous equation model of electoral choice." American Political Science Review 73:1055–70.
McClosky, Herbert
1958 "Conservatism and personality." American Political Science Review 52:27–45.
McKelvey, Richard D., and William Zavoina
1975 "A statistical model for the analysis of ordinal level dependent variables." Journal of Mathematical Sociology 4:103–20.
Miller, George
1956 "The magical number seven, plus or minus two." Psychological Review 63:81–89.
Natchez, Peter B., and Irvin C. Bupp
1970 "Candidates, issues, and voters." In Edward C. Dreyer (ed.), Political Opinion and Behavior. Belmont, CA: Wadsworth.
Repass, David E.
1971 "Issue salience and party choice." American Political Science Review 65:389–400.
Stokes, Donald E.
1966 "Spatial models of party competition." In Angus Campbell, et al. (eds.), Elections and the Political Order. New York: John Wiley and Sons.
Sullivan, John L., James E. Pierson, and George E. Marcus
1978 "Ideological constraint in the mass public: a methodological critique and some new findings." American Journal of Political Science 22:233–49.
Zajonc, Robert B.
1971 "The concepts of balance, congruity, and dissonance." In Peter Suedfeld (ed.), Attitude Change: The Competing Views. Chicago: Aldire-Atherton.

Does Time of Voting Affect Exit Poll Results?

RONALD J. BUSCH AND JOEL A. LIESKE

DESPITE a vast and growing literature on American voting behavior, little is known about the times when different groups of people vote. And the few studies which have systematically analyzed the timing of the vote are inconclusive (Fuchs and Becker, 1968; Mendelsohn and Crespi, 1970; Klorman, 1976). While early studies suggest significant differences in the times people vote, some recent research finds at most, scant differences. Thus, Klorman (1976) concludes that the major situational factors that might be expected to condition the time of voting—employment status, occupation, age, religion, and income—largely fail to do so. However, his conclusions are based on aggregated national data that may tend to cancel or minimize time of voting effects in different metropolitan areas and regions of the country.

Scholarly interest in the timing issue, thus far, has been muted by the paucity of temporal voting data, a lack of concern for election day voting constraints, and the normative and political ramifications of an early election "call."[1] However, the growing use of exit polls by the media and

---

[1] There is some evidence that media projections of a Reagan landslide in the 1980 presidential election may have reduced voting turnout on the West Coast by as much as 6 to 11 percent (Jackson and McGee, 1981; Robinson and Sheehan, 1983).

Abstract This study analyzes time of voting effects on the results of exit poll data collected in a Cleveland municipal tax referendum. The data reveal significant differences both in the times different groups vote and in the reported levels of support for the tax levy by time of day. The temporal variation in the reported levels of support for the tax levy are then shown to be directly related to the times when different support groups cast their ballots. Collectively, these findings suggest the need for day-long interviewing in the design of exit polls.

Ronald J. Busch and Joel A. Lieske are Associate Professors in the Department of Political Science, Cleveland State University.

Public Opinion Quarterly Vol. 49:94–104 © by the Trustees of Columbia University
Published by Elsevier Science Publishing Co., Inc. 0033-362X/85/0049-94/$2.50

polling organizations to project and analyze American elections both allows and now requires a systematic reexamination of this issue.[2]

In a post-1980 election interview, George Gallup suggested (Sanoff, 1981:32) that "by using exit polls, it would be fairly simple to call the national election by 11:00 A.M. on election day except in the very closest election." To the contrary, our analysis of the time of voting will demonstrate the risk of considerable sampling bias, especially when interviewing is confined to a three- or four-hour period on election day. Moreover, our results will show that these temporal sampling effects can be largely explained by the differences in times when different types of voters cast their ballots.

Data and Method

Our data are drawn from an exit poll of 547 Cleveland respondents who voted in a special election on February 17, 1981, for a proposed increase of one-half percent in the city's income tax.[3] We had two reasons for choosing to study a special municipal tax referendum. First, the sheer number of municipalities and the legal and constitutional requirements of citizen approval of local revenue issues make municipal elections the most frequent type in American politics. Second, tax referenda represent the most difficult case to predict and analyze. Why is this so? Relative to the more predictable partisan patterns that prevail in general elections at the national and state levels, tax referenda are generally unstructured events, often characterized by low levels of political information and interest. As a result, they tend to be low turnout affairs with large numbers of undecided voters.

To conduct the exit poll, we employed a two-stage cluster design (Mendenhall, et al., 1971). The first stage consists of a systematic random sample of 25 precincts selected from a total frame of 645 precincts in the city.[4] The second stage consists of 24 exit interviews conducted at selected time points during polling hours (6:30 A.M. to 7:30 P.M.). Thus, a total of 600 interviews were targeted for election day. The exit poll yielded 547 useable questionnaries, for an interview rate of 91 percent. Field problems accounted for the loss of 53 scheduled interviews.[5]

---

[2] Though the Political Unit of CBS News began experimenting with exit polls as early as 1967, their use on a massive scale is a relatively recent development (Plissner and Mitofsky, 1982). In the 1980 presidential election, each of the three major networks, in collaboration with the nation's leading newspapers and wire services, conducted election day exit polls. Surveying some 36,000 actual voters, these polls were the principal method employed by the media to analyze the characteristics and voting decisions of the American electorate (Levy, 1983: 54). The ability of NBC News to scoop its competitors in projecting the 1980 presidential election suggests that exit polls have come of age and have been incorporated into the rapidly changing technology of political polling.

[3] These data were made available by WKYC-TV, the Cleveland NBC affiliate, for whom one of the authors serves as elections consultant and pollster. We wish to thank the station management and Mr. Clfford Abromats, now News Director for WABC-New York, for their support.

The 24 interview time points reflect our expectations about peak turnout periods and the times different groups of voters would be most likely to cast their ballots. We hypothesized, for instance, that housewives and retirees would be more likely to vote during the midmorning and midafternoon hours in order to avoid the queues of voters often found in the early morning, at noon, and in the late afternoon. Conversely, we expected professional and white-collar workers to cast their ballots in the early morning and late afternoon, while we thought factory workers would turn out in greater numbers during the late afternoon and early evening hours.

In order to ensure a representative sample, we clustered 19 of the 24 scheduled interviews in each precinct within the peak turnout periods of early morning, midday, and late afternoon. Interviewing during the peak periods was conducted at 15- to 20-minute intervals. During the slack hours, interviews were spaced at intervals of 45 to 90 minutes.

Because we expected approximately equal numbers of men and women to turn out in the special election,[6] we instructed our exit interviewers to alternate the sex of the respondent from one interview to the next. In keeping with standard exit polling procedures, our interviewers distributed a one-page questionnaire printed on both sides. To further encourage respondent cooperation, we placed an NBC logo at the top and titled it "Secret Ballot." The "Ballot" contained questions on how the respondent had voted on the tax levy, the reasons for his or her vote, and standard demographic characteristics such as race, sex, age, education, occupation, employment status, and party identification.

Our primary objective was to test the hypothesis of temporal sampling effects, namely, does the timing of election-day interviews significantly alter the representativeness of the sample and the reported levels of support for the tax levy? We hypothesized that levels of support for the tax levy would vary by time of day because they reflect differences in the times when different groups of voters come to the polls. To test this hypothesis, we first analyzed the times different groups voted. Next, we checked to see if the reported levels of support for the tax levy also varied by time of day. Finally, we determined if temporal differences in support for the tax levy were related to the times different groups voted.

---

[4] The sample size of the first stage was largely determined by the amount of money available and prevailing wage rates for exit poll interviewers. A budget line of $1000 and a $40 per day pay scale set a maximum of 25 poll workers. Poll workers were responsible for conducting interviews at the 25 designated exit precincts. One poll worker was responsible for two precincts located in the same polling place.

[5] The refusal rate for this study was 6 percent. The remaining attrition, amounting to 3 percent, was a result of the physical features of the polling place that made scheduled interviews impossible. Voting sites with multiple exits and those in high-rise apartments reduced the number of "exiting" voters available for interviewing at designated stations.

[6] While there are more women registered to vote than men, the higher turnout rate for men in Cleveland usually results in approximately equal numbers of male and female voters.

Results

On the basis of the responses elicited in the exit poll, we obtained a 67.0 percent level of support for the tax levy.[7] With a standard error of 2.95 percent,[8] this estimate was within allowable sampling error of the actual result, 62.2 percent, at the .05 level of confidence.

In Table 1 we present the distribution of exit voters in the Cleveland special election by time of interview for eight selected voter characteristics. The times of interview are grouped into five equal intervals: (1)

---

[7] This estimate was computed using the formula:

$$\hat{p} = \frac{\sum_{i=1}^{n} M_i \hat{p}_i}{\sum_{i=1}^{n} M_i}$$

where $\hat{p}$ is the weighted margin of support for the tax levy; $\hat{p}_i$ is the margin of support for the tax levy among respondents in precinct i; M_i is the total number of voters in precinct i; and n is the number of sample precincts.

[8] The standard error for a two-stage cluster sample is given by the square root of the sampling variance, $V(\hat{p})$, where

$$V(\hat{p}) = \frac{(N-n)}{Nn\bar{M}^2}\,\frac{\sum_{i=1}^{n} M_i^2(\hat{p}_i - \hat{p})^2}{n-1} + \frac{1}{Nn\bar{M}^2}\sum_{i=1}^{n} M_i^2 \frac{(M_i - m_i)}{M_i}\,\frac{\hat{p}_i\hat{q}_i}{m_i - 1},$$

and where $\hat{p}$ is the estimated proportion of voters in the city with a given characteristic, N is the total number of precincts, n is the number of precincts selected in a simple random sample, M_i is the total number of voters who turn out in precinct i, m_i is the number of exit voters interviewed in precinct i.

$\bar{M} = \dfrac{\sum_{i=1}^{N} M_i}{N}$ is the citywide average number of voters in each precinct, and $\hat{p}_i$ is the estimated proportion of voters in precinct i with a given characteristic, $\hat{q}_i = 1 - \hat{p}_i$.

early morning (6:30 A.M. to 9:00 A.M.), (2) midmorning (9:00 A.M. to 11:30 A.M.), (3) noon (11:30 A.M. to 2:00 P.M.), (4) midafternoon (2:00 P.M. to 4:30 P.M.), and (5) late afternoon(4:30 P.M. to 7:00 P.M.).[9] Since

Table 1. Distribution of Exit Voters by Time of Interview[a]

	Time of Interview						
Characteristics	*Early Morn.*	*Mid Morn.*	*Noon*	*Mid After.*	*Late After.*	*N*	*Cramer's V*
Employment							
Not employed	10.0	25.8	38.7	15.9	9.6	200	.35
Employed	18.1	13.0	18.0	18.3	32.6	341	
Sex							
Male	17.9	19.9	30.3	14.2	17.7	270	.19
Female	12.6	15.5	21.7	20.2	30.0	272	
Occupational status							
Unskilled	3.1	12.2	28.4	32.7	23.6	31	.16
Skilled	16.9	13.9	22.4	21.8	25.0	115	
Clerical	17.9	10.4	6.6	16.6	48.6	69	
Managerial	13.8	13.6	26.3	0.0	46.3	17	
Professional	25.9	10.3	21.4	15.6	26.8	83	
Education							
6th grade	11.7	24.7	42.5	0.0	21.1	15	.14
9th grade	21.0	20.9	22.1	16.2	19.8	53	
High school	9.5	20.3	27.1	22.3	20.8	228	
Some post high	13.3	16.4	29.7	15.3	25.3	164	
Coll. grad.	29.5	10.4	14.8	12.3	33.0	85	
Age							
18–24	17.8	13.1	27.4	14.0	27.7	64	.14
25–34	18.3	9.1	21.7	19.4	31.5	124	
35–44	12.7	16.4	24.5	25.4	21.0	79	
45–54	10.7	27.7	13.9	22.4	25.3	86	
55–64	15.0	19.6	28.8	13.3	23.2	112	
65 +	14.7	26.0	39.1	8.7	11.4	79	
Race							
White	16.6	15.5	27.0	16.7	24.3	295	.11
Black	13.0	21.2	23.6	18.7	23.4	244	
Other	8.5	14.0	28.0	7.7	42.2	8	
Party identification							
Democratic	14.8	19.1	24.6	17.5	24.1	372	.07
Republican	14.2	12.2	33.0	19.2	21.4	62	
Independent	19.1	14.6	24.6	17.1	24.6	87	
Other	3.9	21.0	33.3	10.0	31.7	19	
Tax							
Anti-tax	14.5	16.5	19.2	23.6	26.2	173	.14
Pro-tax	15.3	17.7	29.2	14.5	23.3	369	
(N)	(83)	(97)	(139)	(99)	(129)	(547)	

[a] Table entries are in percentages. These may not add up to 100 percent due to rounding.

---

[9] Interviewing was concluded at 7:00 P.M., one half hour before the polls closed. Early termination of the interviewing does not seem to have affected the poll results in an unexpected way. Our estimate of tax support for the late afternoon subsamples, 64.3 percent, is within the margin of allowable sampling error.

time of voting may vary with voter characteristics in nonlinear ways, we have used Cramer's V as our measure of association.[10]

The results show that the only voter characteristic which significantly influences time of voting is employment status. The Cramer's V statistic yields a moderate correlation of .35. Employed people, with fixed hours of work, are more likely to vote in the early morning and late afternoon. By contrast, unemployed voters, a category that includes housewives, students, and retirees, have a more flexible day and thus vote more often in the midmorning and noon periods.

The low Cramer's V correlations for the other voter characteristics show that time of voting is only marginally affected by gender, occupational status, education, age, race, party identification, and position on the tax levy. However, there are generally one or two subgroups within each characteristic whose voting times are different from the others, e.g., college graduates and professionals.

In order to identify the subgroups which are most dissimilar, we computed indices of dissimilarity between all pairwise profiles for each voter characteristic.[11] The results are presented in Table 2. The data show that the most dissimilar subgroups are those with the most discretionary time or irregular work hours: namely, the unemployed, unskilled, grade school educated, college educated, professional, and retired voters.

Up to this point, we have analyzed the times when different groups of people vote (Table 1) and those groups which are most dissimilar from one another in their times of voting (Table 2). We now wish to know whether the levels of tax support also vary at different times of the day and how group differences affect the overall levels of tax support.

To examine these issues, we compared the levels of support for the tax levy by voter characteristic and time of voting. The results are presented in Table 3. As expected, the data show that support for the tax levy varied

---

[10] Because the reported percentages in Table 1 are based on weighted data, it would be misleading to report the chi-square values. The weighting procedure makes all the chi-square values significant. Cramer's V is a nonlinear measure of overall association, derived from the chi-square statistic, that takes on values between 0 and 1.

[11] The index of dissimilarity is a measure of inequality in the cumulative distribution of two variables. With a possible range of values between 0 and 1, it represents the percentage of exit voters who would have to shift voting times for the two distributions to become identical (Duncan and Duncan, 1955). Indices of dissimilarity, D, were computed by the formula:

$$D = \frac{1}{2} \sum_{i=1}^{k} |X_i - Y_i|$$

where X_i, Y_i are the uncumulated proportions of respondents in each category, and k is the total number of categories.

Table 2. Indices of Dissimilarity of Time of Voting for Employment Status, Occupational Status, Education, Race, Age, Sex, Party Identification, and the Tax Levy

Variable Category	*(1)*	*(2)*	*(3)*	*(4)*	*(5)*
Employment status					
(1) Not employed					
(2) Employed	33.5				
Sex					
(1) Male					
(2) Female	18.3				
Occupational status					
(1) Unskilled					
(2) Skilled	16.9				
(3) Clerical	39.8	24.6			
(4) Managerial	34.8	25.2	22.0		
(5) Professional	26.0	10.8	22.8	27.7	
Education					
(1) 6th grade					
(2) 9th grade	25.5				
(3) High school	22.3	12.1			
(4) Some post high	21.1	13.1	10.9		
(5) Coll. grad.	42.0	21.7	32.2	23.9	
Age					
(1) 18–24					
(2) 25–34	9.7				
(3) 35–44	14.7	16.1			
(4) 45–55	23.0	21.6	15.6		
(5) 55–64	8.0	17.6	12.0	19.2	
(6) 65 +	24.6	34.4	26.2	29.2	16.7
Race					
(1) White					
(2) Black	7.8				
(3) Other	19.0	23.2			
Party identification					
(1) Democrat					
(2) Republican	10.2				
(3) Independent	4.8	10.5			
(4) Other	18.2	19.4	22.2		
Tax					
(1) Anti-tax					
(2) Pro-Tax	12.0				

significantly by the time of day. In fact, the range of support varies by more than 20 percent, from a noon high of 75.5 percent to a mid-afternoon low of 55.4 percent. If interviewing had occurred only during the mid-afternoon, our estimate of tax support would have understated the actual level of support by almost 7 percent. If interviewing had occurred only during the noon hours, our estimate would have overstated the actual level of support by more than 13 percent.

Second, despite remarkable levels of broad-based support for the February 17 tax levy, the data reveal considerable group differences. The

Table 3. Support for Tax Levy, by Time of Voting[a]

Characteristics	*Time of Voting* *Early Morn.*	*Mid Morn.*	*Noon*	*Mid After.*	*Late After.*	*Total Sample*
Employment status						
Not employed	89.2	67.9	84.1	66.5	68.4	76.0
Employed	61.8	68.1	69.4	50.1	63.7	62.4
Sex						
Male	65.7	61.4	80.2	51.0	61.9	66.4
Female	72.0	77.1	69.2	57.7	67.1	67.8
Occupational status						
Unskilled	100.0	47.3	54.2	78.6	47.4	65.9
Skilled	63.3	58.5	79.9	34.3	50.1	57.4
Clerical	71.4	92.2	0.0	44.0	50.0	55.2
Managerial	100.0	67.2	60.7	—[b]	37.3	77.6
Professional	61.9	62.0	72.7	78.9	44.6	69.8
Education						
6th grade	100.0	100.0	100.0	—[b]	50.1	89.4
9th grade	92.1	70.3	92.1	51.4	75.2	77.3
High school	53.7	57.1	67.4	52.6	60.[illegible]	59.2
Some post high	73.8	82.9	78.0	49.3	60.8	69.4
Coll. grad.	64.0	64.4	79.3	85.5	73.0	72.0
Age						
18–24	60.0	73.6	89.3	58.8	70.4	72.4
25–34	78.9	73.8	54.6	46.5	51.8	58.4
35–44	72.0	71.6	53.8	48.2	59.0	58.7
45–54	33.4	70.6	100.0	74.8	78.4	73.6
55–64	60.4	51.0	81.6	43.8	59.4	62.2
65 +	88.2	77.4	94.5	79.9	100.0	88.4
Race						
White	65.6	57.4	69.2	47.1	62.3	61.4
Black	75.6	83.8	86.2	66.6	70.1	76.8
Other	0.0	0.0	100.0	100.0	33.6	49.4
Party identification						
Democratic	71.7	71.7	71.7	56.4	68.2	68.0
Republican	66.2	68.8	77.4	75.5	64.0	71.5
Independent	55.6	46.6	80.6	40.8	59.0	59.0
Other	100.0	78.6	100.0	0.0	27.1	62.4
Overall Support	68.3	68.5	75.5	55.4	64.3	67.0
(N)	(83)	(97)	(139)	(99)	(129)	(547)

[a] Table entries are in percentages. [b] No respondents of this type.

greatest levels of support for the levy, for instance, were provided by the unemployed, blacks,and the elderly, groups least likely to bear the burden of the tax increase. By comparison, much lower levels of support for the tax levy are discernible among the groups most likely to incur the costs of the increase: the employed, whites, and those aged 25–44. Thus, the level of support among the unemployed is 13.6 percent higher than for the employed. The level of support among blacks is 15.4 percent higher than among whites. And the level of support among voters over 65 is, on the average, about 30 percent higher than voters aged 25–44.

Finally, it is clear from the curvilinear data for occupation and education that support for the tax levy was also conditioned by the emergence of a classic "top-bottom" voting coalition. Although majorites in each occupational and educational stratum voted in support of the tax increase, levels of support were greater among unskilled, managerial, and professional workers than among skilled and clerical workers. Similarly, support for the tax levy was greater among those without a high school education and those with some post-high school and college training than among the high school educated.

These patterns are consistent, indeed expected, with those observed between support for the tax levy and voter attitudes toward Cleveland's white Republican mayor, George Voinovich; the city's black Democratic council president, George Forbes; and the city's white Democractic former mayor, Dennis Kucinich. Voinovich and Forbes had strongly endorsed the levy and both campaigned for voter approval, the mayor strenuously, on the grounds that the tax increase was necessary to maintain public services. The only opposition to the tax levy came from Kucinich and a small number of neopopulist supporters on city council.

To assess potential constituent effects, we computed the point biserial correlation coefficients between a dichotomous support measure and a ten-point thermometer scale that measured negative versus positive orientations toward the three local politicians.[12] As expected, support for the tax levy is strongly correlated with favorable attitudes toward Voinovich (.57) and Forbes (.37) and unfavorable attitudes toward Kucinich (−.45). Thus, the greater levels of support elicited from white, upper status and black, lower status voters may be respectively linked to the political constituencies of Voinovich and Forbes; while the lower levels of support from white, lower-middle status voters may be associated with the neopopulist constituency of Kucinich.

In summary, then, these data support the hypothesis of temporal sampling effects. That is, levels of support for the tax levy vary by time of day because they reflect differences in the times different support groups come to the polls.

Clearly, the outcome of the election was predictable on the basis of vote returns tabulated during the early morning, midmorning, and midday polling hours. But, the significantly lower levels of support registered in the mid- and late-afternoon hours suggest that, in a closer election, early projections based on morning-only interviewing may entail considerable risk. Only day-long interviewing will yield an accurate estimate that reflects the differential turnout patterns of different voting groups and the consequent temporal differences in the amount of support for a candidate or issue.

---

[12] The question on voter attitudes toward the politicians was worded as follows: "On a scale of 1 to 10 (1 = very negative; 5 = no opinion; 10 = very positive), how do you feel about the following politicians in the city—George Voinovich, George Forbes, and Dennis Kucinich?" As in the rest of this analysis, the responses were weighted by the ratio of the total number of voters in each precinct to the number of exit voters interviewed in each precinct.

The usual technique for controlling temporal sampling effects is fairly simple and easy to implement. One only needs to employ precinct-specific skip intervals that will yield the required number of interviews within each sampled precinct. These skip intervals are computed by dividing the estimated turnout in each precinct by the required number of interviews.

Conclusions

The major purpose of this study was to analyze temporal sampling effects on election outcomes. Our interest in this issue stems, first, from professional forecasting needs, and second, from our desire to understand election day voting constraints. To assess this issue, we analyzed exit poll data collected in a special tax election in Cleveland, Ohio. The results support several conclusions.

First, they show that the times different voting groups come to the polls are significantly affected by employment status. Second, the results suggest that employment status itself may be largely responsible for explaining other temporal differences in group voting. Thus, the groups which deviate most from "normal" or job-centered voting patterns are those with the most discretion in their use of time, i.e., retirees, the unemployed, and professional workers.

Third, our findings suggest that pollsters need to exercise care in setting the times when exit poll interviews are to be conducted. There are two ways in which the timing of interviews can affect exit poll estimates. First, different groups of people vote at different times on election day. Second, levels of support for an issue or candidate depend on the voting preferences of different voting groups. Consequently, sampling frames that limit the interviewing to mornings or afternoons only are likely to produce an unrepresentative cross-section of the voting electorate and therefore unreliable estimates of election outcomes. To minimize this source of error, we strongly recommend day-long interviewing and the strict use of precinct-specific skip intervals to yield a set number of interviews within each sampled precinct.

We believe that exit polls represent a promising new technique for projecting and explaining American voting behavior. Although the temporal effects reported are based on data from a special tax election, we believe that additional polls conducted in different electoral settings and for different types of elections will support these conclusions. Moreover, we believe that scholars would do well to consider using exit polls for more scientific analyses. To date, the principal users of the exit poll have been attracted to it because it allows them to make an early election "call." But the exit poll has other decided advantages. By focusing exclusively on actual voters, it avoids the measurement problems of pre- and postelection surveys and the ecological pitfalls of aggregate data analysis.[13] If it is carefuly executed, it can provide valuable new information on the characteristics and behavior of the American voter.

References

Duncan, Otis, and Beverly Duncan
1955 "A methodological analysis of segregation indices." American Sociological Review 20:210–17.

Fuchs, Douglas, and Jules Becker
1968 "A brief report on the time of day when people vote." Public Opinion Quarterly 32:437–40.

Gallup, George
1951 "The Gallup Poll and the 1950 elections." Public Opinion Quarterly 15:16–22.

Klorman, Ricardo
1976 "Chronopolitics: what time do people vote?" Public Opinion Quarterly 40: 182–93.

Jackson, John E., and William H. McGee III
1981 "Election reporting and voter turnout." Unpublished manuscript.

Levy, Mark R.
1983 "The methodology and performance of election day polls." Public Opinion Quarterly 47:54–67.

Mendelsohn, Harold, and Irving Crespi
1970 Polls, Television, and the New Politics. Scranton, PA: Chandler.

Mendenhall, William, Lymann Ott, and Richard L. Scheaffer
1971 Elementary Survey Sampling. Belmont, CA: Wadsworth.

Miller, Mungo
1952 "The Waukegan study of voter turnout prediction." Public Opinion Quarterly 16:381–98.

Plissner, Martin, and Warren Mitofsky
1982 "Voting twice on election day." Public Opinion (August/September):14–16.

Robinson, Michael J., and Margaret A. Sheehan
1983 Over the Wire and on TV: CBS and UPI in Campaign '80. New York: Russell Sage Foundation.

Sanoff, Alvin
1981 "The perils of polling 1980: how the takers of the nation's pulse missed the beat." Washington Journalism Review, Jan/Feb 1981.

---

[13] Without the voting histories of respondents, it is difficult to identify likely voters in preelection surveys (Gallup, 1951; Miller, 1952). In addition, preelection surveys suffer from voter indecision and last minute shifts in voter sentiment. Postelection surveys, by comparison, are often contaminated by the problems respondents have in recalling their voting decisions and by postelection rationalization. Finally, precinct vote results provide no explicit information about voter attitudes, social characteristics, and partisan orientations other than those which can be inferred from ecological analysis of aggregated data.

The Importance of Replicating a Failure to Replicate: Order Effects on Abortion Items

GEORGE F. BISHOP,
ROBERT W. OLDENDICK, AND ALFRED J. TUCHFARBER

THE EXPECTATION that all experiments should be replicated has become a hallowed rule of evidence in scientific research. Yet far too few experiments are ever replicated. Here we will describe a failure to replicate one such study: Schuman et al. (1981), "Context effects on survey responses to questions about abortion." We will also show, however, that it is just as important to replicate such failures to replicate, not only because it is necessary to demonstrate that the failure, itself, is a reliable one; but also because, unlike a successful replication, a failure may occur for various reasons (see Aronson and Carlsmith, 1968: 20–22).

In this case, moreover, we did not expect a failure. Indeed, we had borrowed the experiment on the abortion issue by Schuman and his colleagues because it provided a good example of a sizable, well-

Abstract A failure to replicate a previously demonstrated order effect on responses to questions about abortion led to a second experiment which was designed to test the hypothesis that the failure was due to a change in the question context which had immediately preceded the abortion items in the original experiments by Schuman, et al. Unlike the results of the first experiment, the results of this second experiment showed that the order effect reported by Schuman and his associates does replicate. In addition, the data indicate that context may condition the magnitude of this order effect. The authors also suggest that the effect may be due to the implicit *contrast* between the reasons given for abortion by the items when they are asked in one order rather than the other.

George F. Bishop is Associate Professor of Political Science and a Senior Research Associate in the Institute for Policy Research at the University of Cincinnati. Robert W. Oldendick is Assistant Director and Alfred J. Tuchfarber is Director of the Institute for Policy Research at the University of Cincinnati. The research reported in this paper was supported by a grant from the National Science Foundation (SES81-11404). The authors want to thank Pat Brown for her assistance in processing the data, as well as her comments on our interpretation of the results. The authors also want to thank the anonymous reviewers for their helpful criticisms and suggestions.

Public Opinion Quarterly Vol. 49:105–114 © by the Trustees of Columbia University
Published by Elsevier Science Publishing Co., Inc. 0033-362X/85/0049-105/$2.50

demonstrated order effect, having been replicated at least twice in national surveys. Thus, we thought it would be useful for testing the conventional wisdom that such effects can be eliminated by interposing a "buffer" of questions on unrelated topics between items that are known, or suspected, to affect one another. In other words, we had assumed that the order effect reported by Schuman and his associates would replicate when we repeated the conditions of their split-ballot experiment, but would diminish or disappear entirely when a buffer was used in a third experimental condition (cf. Bishop et al., 1984; Schuman et al., 1983).

Experiment I

This first split-ballot experiment was done in November and December of 1981 as part of the Greater Cincinnati Survey (GCS), an omnibus, random-digit-dialed (RDD) telephone survey conducted by the Behavioral Sciences Laboratory at the University of Cincinnati.[1] It consisted of three conditions to which respondents were randomly assigned:

Form Q. Respondents in this group were asked the two questions about the abortion issue, used by Schuman et al. (1981), in the following sequence:

1. Do you think it should be possible for a pregnant woman to obtain a *legal* abortion if there is a strong chance of serious defect in the baby? (This question is referred to below as the *birth defect* item.)
2. Do you think it should be possible for a pregnant woman to obtain a *legal* abortion if she is married and does not want any more children? (This question is referred to below as the *woman's right* item.)

Form R. Here respondents were asked the same two questions, but in reverse order—i.e., the woman's right item, first, and the birth defect item, second.[2]

Form S. In this condition respondents were asked the birth defect question, first, as in Form Q. They were not, however, asked the woman's right question immediately afterwards, but rather after being exposed to a buffer of approximately 40 questions on topics unrelated to the abortion issue.

---

[1] The response rate (fully completed interviews) for this survey was 74.0 percent. The refusal rate was 14.4 percent. The remainder consisted of partially completed interviews (3.6 percent), potential interviews that were not completed because of a language barrier, a hearing problem, illiteracy, senility, or physical illness (6.5 percent), or because the selected respondent was away on vacation, a business trip, or some other matter (1.5 percent).

[2] Schuman et al., (1981) have referred to the woman's right item as the *general* abortion item and to the birth defect item as the *specific* abortion item. Because the terms *general* and *specific* may be somewhat vague, as well as misleading, we have adopted the recommendation of one of the anonymous reviewers to use a more descriptive set of terms.

Furthermore, none of the questions which preceded these three forms in the interview dealt in any way with the topic of abortion. So there were no apparent contaminants in the research design.

Table 1 shows the data from both our study (GCS-1981) and the two SRC studies by Schuman and his associates which have been combined here to conserve space.[3] In their experiments, the order in which the

Table 1. Order Effects on Responses to Abortion Items, by Survey

I. Surveys: SRC-1979 June and August (combined)

Order: Defect/Right

Birtht Defect item	Woman's Right item YES	Woman's Right item NO	
Yes	47.9	37.4	85.3
No	0.8	13.8	14.6
	48.7	51.2	99.9% (593)

Gamma = .91

Order: Right/Defect

Birth Defect item	Woman's Right item YES	Woman's Right item NO	
Yes	61.4	24.2	85.6
No	2.9	11.5	14.4
	64.3	35.7	100.0% (624)

Gamma = .83

Woman's right item marginals by order: $\chi 2 = 29.97$, df = 1, $p < .001$
Birth defect item marginals by order: $\chi 2 = 0.02$, df = 1, $p > .50$
Interaction (response to *woman's right* item by response to *birth defect* item by order): $\chi 2 = 1.93$, df = 1, $p = .17$.

II. Survey: Greater Cincinnati—1981 November/December

(Form Q)
Order: Defect/Right

Birth Defect item	Woman's Right item YES	Woman's Right item NO	
Yes	46.2	31.7	77.9
No	0.6	21.6	22.2
	46.8	53.3	100.1% (357)

Gamma = .96

(Form [illegible])
Order: Right/Defect

Birth Defect item	Woman's Right item YES	Woman's Right item NO	
Yes	49.5	25.1	74.6
No	1.9	23.5	25.4
	51.4	48.6	100.0% (362)

Gamma = .92

Woman's right item marginals by order: $\chi 2 = 1.49$, df = 1, $p = .22$
Birth defect item marginals by order: $\chi 2 = 0.82$, df = 1, $p = .37$
Interaction (response to *woman's right* item by response to *birth defect item by order*): $\chi 2 = 1.18$, df = 1, $p = .28$

---

[3] There was no significant difference between the two SRC surveys in the magnitude of the order effect on the woman's right item (response to the woman's right item by order by SRC survey: $\chi^2 = 1.22$, df = 1, $p = .27$). The data for the SRC-1979 June survey were obtained from Table 2 in Schuman et al. (1981). Those for the SRC-1979 August survey were made available by Stanley Presser, whom the authors would like to thank. We alone of course are responsible for any errors in the analysis and interpretation of these data.

questions were asked made a substantial difference in the percentage of respondents who said yes to the woman's right item: when it was asked first, people were much more likely to say yes (64.3 percent) than when it was asked after the birth defect item (48.7 percent). In our experiment, however, the order of the questions made a small, but statistically insignificant difference in the results for the woman's right item: when it was asked first, respondents were only slightly more likely to say yes (51.4 percent) than when it was asked after the birth defect item (46.8 percent).[4]

Some readers might argue that the order effect is still in the same direction as that found in the SRC experiments, but that it is not large enough to be statistically significant with the sample size used. A more appropriate indication of a failure to replicate,then, would be to show that the magnitude of the order effect in our experiment (4.6 percent) differs from that found in the SRC surveys (15.6 percent). That, in fact, is what we found: a significance test of the three-way interaction—response to the woman's right item by order by survey (SRC June and August, 1979 vs. GCS 1981)—shows that the magnitude of the order effect in the SRC studies was significantly larger than that observed in our experiment ($\chi^2 = 5.70$, df = 1, $p < .02$).[5] The generalizability of the order effect discovered by Schuman and his associates would thus seem to be questionable.

As expected, question order had virtually no effect on responses to the birth defect item in either the SRC surveys or ours. We also found, as they did, that there were no significant variations in the magnitude of association between the woman's right and birth defect items by form (cf. Table 2).

What accounts for the difference, then, between our findings on the woman's right item and theirs? Random error is always a possible explanation. We also thought it might be due to the fact that our experiment was done at a different time, in a different location, and by a different organization. Yet we knew we had been able to replicate other question form experiments by Schuman and his colleagues in the same metropolitan area at various times. So this hypothesis did not seem plausible.

---

[4] The results for Form S were nearly identical to those shown for Form Q (data not shown here), indicating that the buffer had no effect.

[5] One reviewer pointed out that this chi-squared value was not particularly large and that the test has made no allowance for effects of clustering in the samples. It should be noted, however, that the samples in all of our experiments are based on *simple* random digit dialing (RDD) and that there are therefore no clustering effects of which we are aware. The SRC surveys were also based on random-digit-dialed telephone surveys, and while there may be some clustering effects due to our pooling of the samples, we know of no professionally agreed upon procedure for adjusting for these effects in RDD telephone surveys. Furthermore, even if we were to reduce the obtained chi-squared value by say 30 percent—a rather radical reduction—the reduced chi-squared value of 3.99 would still be statistically significant at the .05 level. The difference between the SRC surveys and ours, then, does not appear to be due to chance (see also footnote 12).

But one other hypothesis did: a difference in the context which immediately preceded the questions about abortion. In the first of the two experiments by Schuman et al., the question immediately preceding the abortion items had to do with labor unions, whereas in their second study, it concerned beliefs about the Soviet Union. Neither of these topics would appear to have any direct implications for a respondent's opinion on the issue of abortion. But the questions which immediately preceded the abortion items in our experiment may have had such implications, albeit not that direct. These items, which were also part of an experiment on order effects and the influence of "buffering," asked respondents about two things:[6]

1. Generally speaking, do you think that *most* people who receive money from welfare could get along without it if they really tried, or do you think that *most* of them really need it?[7]
2. We hear a lot of talk these days about liberals and conservatives. What about yourself—that is, in politics, do you *generally* think of yourself as a liberal, a moderate, or a conservative?[8]

Though neither of these questions seemed to be directly relevant to a respondent's position on the abortion issue, one could argue that they made some respondents, at least, think that the questions on abortion were another way of asking them how "liberal" or "conservative" they were. This context, in other words, may have changed the meaning of the woman's right item in a way that the questions by Schuman et al. about labor unions and the Soviet Union did not. And that is what we tested in a second experiment.

Experiment II

We tested the context hypothesis in June and July of 1982, using another omnibus, RDD telephone survey of the Cincinnati metropolitan area.[9] In this experiment, we randomly assigned respondents to one of four conditions:

---

[6] One randomized subgroup of respondents was asked the welfare item first, and the liberal/conservative item second. Another subgroup was asked the same questions in reverse sequence. And a third group was asked the welfare item first, a buffer of approximately 40 items, and then the liberal/conservative item. The latter item, therefore, immediately preceded the questions about abortion for just two-thirds of the respondents. In Experiment II (see below), this added variation was not included.

[7] Adapted from a question used by the CBS News/New York Times Poll, July 1977.

[8] Adapted from a question used by the Center for Political Studies in the 1978 American National Election Study (see Volume II, Notes, Questionnaire, and Frequencies Addendum, Second ICPSR Edition, Ann Arbor: University of Michigan, 1979).

Form Q. As in our first experiment, these respondents were asked the birth defect item first, and the woman's right item second.

Form R. Similarly, these respondents were asked, as in Experiment I, the woman's right item first and the birth defect item second.

The questions in both Forms Q and R were also asked, as in Experiment I, immediately *after* the questions on welfare beliefs and liberal/conservative self-identification.

Forms QQ and RR. These forms were identical with Forms Q and R, respectively, except that the questions were asked immediately *before* the items concerning welfare beliefs and liberal/conservative self-identification.

In addition, all respondents who answered yes or no to the woman's right item were asked the following open-ended question:

> Why do you think it (should/should *not*) be possible for a pregnant woman to obtain a *legal* abortion if she is married and does not want any more children?

We added this question not only to determine the reasons why people approve or disapprove of abortion under those circumstances, but also to gain some insight into why order effects occur with such items, if and when they do occur.[10]

The data in Table 2 demonstrate the importance of replicating a failure to replicate. For unlike our first experiment, in which we failed to replicate the results of the SRC surveys, we found that the order in which the questions were asked made a significant difference in the percentage of respondents who said yes to the woman's right item. When it was asked first, as in Forms R and RR, the percentage of people saying yes was substantially higher than when it was asked second, as in Forms Q and QQ.[11]

---

[9] The response rate (fully completed interviews) for this survey was 72.6 percent. The refusal rate was 15.4 percent. The remainder consisted of partially completed interviews (3.3 percent), potential interviews that were not completed because of a language barrier, a hearing problem, illiteracy, senility, or physical illness (5.1 percent), or because the selected respondent was away on vacation, a business trip, or some other matter (4.0 percent).

[10] Since this question immediately followed the woman's right items (i.e., came between the woman's right and birth defect items when the former question was asked first), the experiment does not represent an exact replication, and there was some concern that the marginals on the birth defect item may have been affected. As the results to be presented later demonstrate, however, such effects did not occur.

[11] Using our criterion for failure to replicate, the reader might ask how we would explain why the difference between Forms Q and R on the woman's right item was statistically significant in the second experiment (Context A), but not in the first. A significance test of the three-way interaction—response to the woman's right item by Form (Q vs. R) by

Table 2. Order and Context Effects on Responses to Abortion Items[a]

Context A: Asked *after* welfare/libcon questions

(Form Q)
Order: Defect/Right
Woman's Right item

		YES	NO	
Birth Defect item	Yes	43.0	29.6	72.6
	No	0.4	27.0	27.4
		43.4	56.6	100.0% (270)

Gamma = .98

(Form R)
Order: Right/Defect
Woman's Right item

		YES	NO	
Birth Defect item	Yes	49.8	21.7	71.5
	No	4.6	23.9	28.5
		54.4	45.6	100.0% (263)

Gamma = .85

Context B: Asked *before* welfare/libcon questions

(Form QQ)
Order: Defect/Right
Woman's Right item

		YES	NO	
Birth Defect item	Yes	39.2	35.4	74.6
	No	0.4	25.0	25.4
		39.6	60.4	100.0% (260)

Gamma = .97

(Form RR)
Order: Right/Defect
Woman's Right item

		YES	NO	
Birth Defect item	Yes	56.4	18.4	74.8
	No	3.2	22.0	25.2
		59.6	40.4	100.0% (250)

Gamma = .91

Woman's right item marginals by form in Context A: $\chi^2 = 5.38$, df = 1, $P < .05$
Woman's right item marginals by form, in Context B: $\chi^2 = 17.47$, df = 1, $p < .0001$
Birth Defect item marginals by form, in Context A: $\chi^2 = 0.73$, df = 1, $p > .50$
Birth defect item marginals by form, in Context B: $\chi^2 = 0.00$, df = 1, $p > .50$
Three-way interaction (response to *woman's right* item by response to *birth defect* item by form), in Context A: $\chi^2 = 6.60$, df = 1, $p < .05$
Three-way interaction (response to *woman's right* item by response to *birth defect* item by form), in Context B: $\chi^2 = 1.62$, df = 1, $p = .203$

[a] Carried out in the Greater Cincinnati Survey, June/July 1982. Percentages are based on grand totals.

But, as suspected, the broader context in which the abortion questions were asked appears to condition the order effect. When they were asked *before* the questions about welfare beliefs and liberal/conservative identification, the order effect was noticeably larger (59.6 percent said yes in Form QQ vs. 39.6 percent in Form RR) than when they were asked—as in the Fall 1981 experiment—*after* those items (54.4 percent said yes in

---

experiment (I vs. II)—shows, however, that the magnitude of the order effect with Forms Q and R in Experiment II was not significantly larger than that observed in Experiment I ($\chi^2 = 1.28$, df = 1, $p = .26$). So the most plausible explanation is that the apparent difference between Experiments I and II on Forms Q and R is simply due to chance.

Form R vs. 43.4 percent in Form Q). This three-way interaction (response to the woman's right item by form by context) did not, however, reach statistical significance ($\chi^2 = 1.85$, DF = 1, $p = .17$).[12] So the difference in context does not necessarily explain the original failure to replicate in the Fall 1981 experiment. The results may be due to chance, and it would be useful to replicate this experiment with a larger sample.[13]

The reader will notice one other replication in Table 2: question order had no significant effect on responses to the birth defect item in either Context A or Context B, which confirms the findings from our first experiment, as well as those from the SRC surveys. This suggests that it is not as susceptible to order effects as the woman's right item, perhaps because a birth defect is a more socially acceptable reason for having an abortion. That is, the more acceptable the reason, the less susceptible an item giving that rationale may be to such order effects, as compared to items giving less acceptable rationales.

A more important question is how to interpret the order effect on the woman's right item. Schuman and his associates have theorized that some respondents may say yes to the woman's right item when it is asked first because they have a specific reason in mind such as a chance of defect in the child. Furthermore, they argue, when the question about abortion in the case of a birth defect is asked first, this may suggest to respondents that the woman's right item which follows does not pertain to such circumstances. Respondents may therefore feel freer to say no to the woman's right item after having supported, and *subtracted*, the more specific reason about a possible defect.

---

[12] Because there is no significant difference in the magnitude of the order effect between Experiments I and II (see note 11), we can combine the data for Forms Q and R in these two experiments. Doing that, we find that the magnitude of the order effect is significantly larger for forms QQ and RR than it is for Forms Q and R ($\chi^2 = 5.89$, df = 1, $p < .02$), which indicates that the context hypothesis is certainly a plausible one.

Following this argument, the reader might wonder about the comparison of the order effect between the SRC surveys and our surveys if we combine the data from Experiments I and II for Forms Q and R. We made this comparison and found that the magnitude of the order effect on the woman's right item in the SRC surveys (15.6 percent) was still significantly larger than in our surveys (7.3 percent): χ^2 for response by form by survey (SRC vs. GCS) = 4.43, df = 1, $p < .04$. Furthermore, as would be expected, we found no significant difference in the magnitude of the order effect on the woman's right item in the SRC surveys (15.6 percent) and that observed for Forms QQ and RR in Experiment II (20.0 percent): $\chi^2 = 0.65$, df = 1, $p = .421$. Though none of this evidence is conclusive, it does suggest that the magnitude of the order effect for these two birth defect items may be contingent on the broader context in which they are asked.

[13] The reader might also have wondered about why the percentage of "yes" responses to the birth defect item was generally higher (by about 10 percent) in the SRC surveys than in our Greater Cincinnati surveys. This may be due to true population differences. But this difference does not appear to apply to the results for the woman's right item. We have no substantive explanation for this difference. Measurement error may be somewhat greater on the latter item, and this lower reliability would make it more difficult to detect true population differences.

To test the hypothesis of Schuman et al. we asked respondents why they favored or opposed allowing a married woman to obtain a legal abortion if she does not want any more children. Not a single respondent mentioned the chance of a birth defect as the reason why he or she favored allowing an abortion under such circumstances (data not shown here), which should not be surprising given the distinct content of the item. So there is no direct evidence that respondents have the birth defect reason in mind when they say yes to the woman's right item when the latter is asked first.

In fact, the great majority who say yes to the item give as a reason the woman's right rationale that is implicit in the question (60.3 percent), or some variation on that theme (data not shown here). Therein, we think, lies the key to understanding how the order effect occurs. When asked first—that is, by itself—the rationale given in the woman's right item (a married woman who does not want any more children) appears to be the only game in town; it is not, in other words, compared with any other reasons for having an abortion. Hence, it is easy for many respondents to say yes to the proposition, and then give as the reason the freedom-of-choice rationale that is implicit in the question. When, however, they are asked about it after having answered the question about allowing an abortion if there is "a *strong* chance of *serious* defect in the baby" [emphasis added], the freedom-of-choice rationale implicit in the woman's right item does not seem to be as good a reason for allowing an abortion as is a potential birth defect. This implicit contrast between the rationales given in the two items may thus explain the order effect.

Conclusion

To the time-honored rule that all experiments should be replicated, we wish to add an equally important one: *all failures to replicate should be replicated too*. Indeed, we have learned that such failures should not be taken at face value. And though the evidence presented here is not conclusive, we think it may be important in many instances to replicate the context which precedes a known order effect such as that shown to affect responses to questions about abortion. For the results suggest that the magnitude of this particular order effect may depend upon the broader question context in which it occurs. We have also suggested that the explanation for this now well-demonstrated order effect may lie in the implicit contrast between the justifications for abortion given by the two questions. Our research, then, represents more than a mere attempt to replicate a failure to replicate; it tells us more about the conditions under which context affects responses to questions about abortion, and perhaps why.

References

Aronson, Elliot, and J. Merrill Carlsmith
1968 "Experimentation in social psychology." Pp. 1–79 in Gardner Lindzey and Elliot Aronson (eds.), The Handbook of Social Psychology (2nd ed.). Vol. 2. Reading, Mass.: Addison-Wesley.

Bishop, George F., Robert W. Oldendick, and Alfred J. Tuchfarber
1984 "What must my interest in politics be if I just told you 'I don't know'?" Public Opinion Quarterly 48:510–19.

Schuman, Howard, Graham Kalton, and Jacob Ludwig
1983 "Context and contiguity in survey questionnaires." Public Opinion Quarterly 47:112–15.

Schuman, Howard, Stanley Presser, and Jacob Ludwig
1981 "Context effects on survey responses to questions about abortion." Public Opinion Quarterly 45:216–23.

The Polls: The European Peace Movement and Deployment of Nuclear Missiles

CONNIE DE BOER

ON DECEMBER 12, 1979, NATO adopted what is called the "dual resolution" for maintaining unity among the NATO allies. It was decided both to station 572 medium-range missiles in Western Europe and to conduct negotiations with the Soviet Union on arms control so as to obviate the need for this deployment. The peace movement, which had gained a large following as early as 1977–78 with its actions against the neutron bomb, grew into a real mass movement after this resolution. In many countries peace demonstrations have been held on an unprecedented scale and directed mainly against the deployment of new nuclear weapons in Europe.

In this Polls section, a selection has been made from many public opinion polls on the stationing of cruise missiles and Pershing II missiles in Europe, on the use of nuclear weapons, and on the peace movement and the anti-nuclear weapons demonstrations. The results of these polls may vary widely, depending on question wording and sequence.

The popularity in Britain of the peace movement and of the activities organized by it decreased during the first few months of 1983. In January–February 1983, 23 percent were completely opposed to the demonstrations against nuclear weapons in Britain and elsewhere in Europe; this percentage had risen to 33 by the end of April 1983 (SOC). In January 1983, 31 percent stated that the protest had made them more favorable to the campaign to prevent the deployment of cruise missiles in Britain. Early in April, after such events as a big anti-nuclear weapons demonstration at Greenham Common (April 1, 1983), this percentage had dropped to 20. The same survey (MORI, Apr. 8–9, 1983) also indicates that 54 percent had an unfavorable opinion of the women anti-nuclear protesters at Greenham Common.

Women are more inclined than men to support the peace movement and its ideas, as evidenced not only by the existence of women's peace camps in Europe but also by the results here presented of opinion polls conducted in Germany,

Connie de Boer is research assistant of the POLLS Archive and the Baschwitz Institute for Collective Behavior and Public Opinion, University of Amsterdam.

Public Opinion Quarterly Vol. 49:119–132 © by the Trustees of Columbia University
Published by Elsevier Science Publishing Co., Inc. 0033-362X/85/0049-119/$2.50

Britain, and the Netherlands.[1] In Germany it is especially young people who are behind the peace movement, but in the Netherlands the differences between the age groups in their response to relevant questions are minimal. Only when questions are asked on such matters as marching in demonstrations do we observe a greater preparedness among younger people to take an active part.

As for the use of nuclear weapons, many Europeans and Japanese feel that it is not acceptable under any circumstances, not even if we are attacked with nuclear weapons (The Atlantic Institute for International Affairs—Louis Harris Poll, 1983). Wide differences in response are noted among the European countries. The percentage of persons holding this view is lowest in Britain (24 percent), and highest in Spain (61 percent). According to this survey, Americans clearly hold opinions that differ from those of the Europeans: 14 percent of the Americans state that nuclear weapons must never be used. Unilateral disarmament is supported by only 4 percent of the population in the United States. In Europe, the percentages in favor of unilateral disarmament vary from 15 and 17 percent in Norway and Great Britain, respectively, to a majority of 55 percent in Spain.

Many Europeans, including Britons, are against deployment of nuclear weapons in their countries. Comparison of the results of a USICA survey conducted in Western Europe in October 1981 shows that if the No Opinion respondents are excluded, a majority in Belgium, the Netherlands, and West Germany is opposed to the deployment of new nuclear missiles. According to this study, advocates and opponents in Great Britain are divided about evenly, but in a survey by MORI, also conducted in October 1981, 59 percent express opposition to deployment.

In Great Britain, responses to a question asked repeatedly in 1981–84 about whether Britain should allow the cruise missiles to be based there show that a majority was opposed to deployment in 1981 and 1982 (MORI, SOC). The percentage of advocates of stationing does increase gradually in the course of 1983, but this appears to be a transient shift in public opinion. According to the most recent figures for 1984, 53 percent continue to be against and 37 percent in favor of stationing. These results may have been influenced by the fact that in 1983 the deployment of the new nuclear missiles was begun in Great Britain.

An international study in January 1983, commissioned by Newsweek, reveals that only the Netherlands had more opponents to deployment than advocates. In the Netherlands, where opinions have regularly been asked on basing nuclear weapons there, the trend of the results indicates that few people considered it a good decision (varying from 18 percent in 1981 to 11 percent in 1983), and the percentage of respondents who though it "deplorable but acceptable" rose in the course of the 1981–83 period. In 1984, when the question was slightly revised to emphasize that the decision to deploy was about to be taken by the government and the majority of parliament, the outcome showed a slight increase in the percentage who considered it a good decision (16 percent).

In Germany and Great Britain, the deployment of nuclear missiles was begun at the end of 1983, but in the Netherlands implementation of the decision to station them has been delayed again and again, and there is still wide opposition to stationing them. Several surveys show that an ample majority of the Dutch expects deployment eventually to be implemented (Interact/Nieuwe Revu, NIPO/Eldevier, NSS), but the polls also show that the population will resist such deployment: 43 percent say they will not acquiesce in deployment and intend to take part in protest actions (NSS/RVD, Jan. 1984). According to a NIPO/Elsevier survey in August 1983, 12 percent consider that if necessary the cruise missiles must be kept away by force from Woensdrecht, the place the government has designated for deployment.

---

[1] See Tom W. Smith, "The polls: gender and violence." Public Opinion Quarterly 48:384–396.

The results here presented are all based on national representative samples of adults, and wherever the data were available, the sample number is given. The date stated with each table is the date or period when the survey was conducted. Where percentages do not add up to 100, the apparent discrepancy is due to rounding. Research results have been selected from the collections of the Polls Archives, University of Amsterdam, Netherlands.

The following organizations have graciously granted permission to use their research materials:

Demoskopie—Institut für Demoskopie, Allensbach, West Germany
Doxa—Instituto per la ricerche statistiche el'analise dell'opinione pubblica, Italy
Elsevier—Elsevier's Magazine, Netherlands
EMNID—EMNID—Institut GmbH & Co., West Germany
Eurodim—Eurodim, Greece
Gallup International
Harris—Louis Harris and Associates, USA
Interact—Interact Ltd., Netherlands
Inter/view—Inter/view Ltd., Netherlands
Lagendijk—Bureau Lagendijk, Netherlands
MORI—Market & Opinion Research International, Great Britain
Newsweek—Newsweek Inc., USA
Nieuwe Revu—Nieuwe Revu, Netherlands
NIPO—Netherlands Institute of Public Opinion, Netherlands
NOP—NOP Market Research Ltd., Great Britain
NSS—N.V. v/h Nederlandse Stichting voor Statistiek, Netherlands
RVD—Rijksvoorlichtingsdienst (Government Information Agency), Netherlands
SINUS—SINUS, West Germany
SOC—Social Surveys (Gallup Poll) Ltd., Great Britain
USICA—United States International Communication Agency, USA

The Peace Movement

USA—Harris: In many countries of the world, people have been demonstrating against the spread of nuclear weapons. In general, are you in sympathy with the anti-nuclear weapons movement or not?

	In Sympathy	*Not in Sympathy*	*Not Sure*
Mar. 5–21, 1982, *n* = 1599 (for Asahi Shimbun)	55%	39%	6%
Apr. 20–25, 1983, *n* = 1253 (for Business Week)	45	48	7

USA—Harris: Many people in Western Europe are worried that if the U.S. puts new nuclear missiles into Western Europe, there will be a nuclear confrontation there between the U.S. and Russia, which could lead to a nuclear war. Do you share this worry of people in Western Europe, or not?

	Share This Worry	*Don't Share It*	*Not Sure*
Mar. 1983	52%	43%	5%
Nov. 9–14, 1983, *n* = 1252	57	39	4

Great Britain—SOC: Demonstrations against nuclear weapons have recently taken place in Britain and elsewhere in Europe. Do you personally feel: completely in agreement with, to some extent in agreement with, to some extent opposed to, completely opposed to these particular demonstrations?

	Completely Agree	*Agree to Some Extent*	*Opposed to Some Extent*	*Completely Opposed*	*Don't Know*
Nov. 4–10, 1981	23%	29%	15%	24%	8%
Jan. 28–Feb. 3, 1983	18	36	17	23	5
Apr. 20–26, 1983	14	31	18	33	4
May 25–30, 1983	12	32	18	33	4

Great Britain—SOC: Which of these statements best describes your views aobut nuclear weapons?

	Sept. 10–15, 1980							*Nov. 5–10*	*Nov. 10–15*
	All	*M*	*F*	*16–24*	*25–44*	*45–64*	*65+*	*1981*	*1982*
Worried and willing to join demonstration against them	7%	7%	7%	10%	8%	6%	3%	10%	9%
Worried and willing to write to MP, newspaper	4	4	5	3	5	5	3	7	7
Worried, but won't do anything about it	17	15	18	17	17	16	19	17	17
Worried but do not think anything can be done about them	37	32	42	32	37	40	41	38	39
All worried	65	58	72	62	67	67	67	72	72
Not worried about them	30	39	23	36	30	29	28	26	25
Don't know	4	3	5	2	3	5	7	3	3

Great Britain—MORI/Daily Star: Regardless of your views on nuclear weapons, do you approve or disapprove of the tactics used by anti-nuclear protesters?

	Approve	*Disapprove*	*No Opinion*
Apr. 8–9, 1983, *n* = 1038	29%	58%	13%

Do you think the anti-nuclear protesters contribute to peace, or do you think that they make peace more difficult to achieve?

	Contribute	*No Difference*	*Make More Difficult*	*Don't Know*
Apr. 8–9, 1983, *n* = 1038	24%	24%	46%	6%

Do you have a favorable or unfavorable opinion of the women anti-nuclear protesters at Greenham Common?

	Favorable	*Unfavorable*	*No Opinion*
Apr. 8–9, 1983, *n* = 1038	26%	54%	20%

Has the protest made you more favorable or less favorable to the campaign to prevent cruise missiles being based in Britain or has it made no difference one way or the other?

	More	*Less*	*No Difference*	*No Opinion*
Jan. 18, 1983, *n* = 1053	31%	14%	53%	2%
Apr. 8–9, 1983, *n* = 1038	20	14	60	4

Great Britain—NOP/Daily Mail: Have the peace women at Greenham Common made you more likely or less likely to think Britain should give up its nuclear weapons?

Nov. 1983, n = 1060	*More Likely*	*Less Likely*	*No Difference*	*Don't Know*
All	17%	14%	68%	7%
Men	15	15	69	1
Women	18	13	67	2
18–34	19	12	67	2
35–54	19	12	66	2
55+	11	17	70	2

West Germany—EMNID: On October 10 there was this big peace demonstration in Bonn. Do you approve of this peace demonstration having been held, or do you disapprove?

Oct. 15–Nov. 9, 1981, n = 1007	*Approve*	*Disapprove*	*No Answer*
All	65%	32%	2%
Men	62	37	1
Women	68	28	3
–19	70	26	4
20–29	77	23	—
30–49	67	31	1
50–64	64	35	1
65+	50	44	6

West Germany—EMNID: There have been protest meetings of different groups and organizations against the deployment of new missiles. Do you personally tend to approve or to disapprove of demonstrations against these missiles?

Nov. 24–Dec. 5, 1983 n = 1002	*Tend to Approve*	*Depends on Nature of Protest*	*Tend to Disapprove*	*Don't Know, Indifferent*	*No Answer*
All	30%	36%	21%	11%	2%
Men	29	36	28	6	1
Women	31	37	15	15	3
–19	46	34	16	3	—
20–29	39	29	16	10	6
30–49	30	37	23	8	1
50–64	21	39	28	11	1
65+	23	39	16	22	—
Attitude to deployment					
Pro (27%)	7	36	52	5	—
Contra (57%)	49	35	8	6	1

West Germany—SINUS (conducted for the Federal Chancellor's Office)

	Oct.	*Sep. 1982, n = 1600*							
	1981	*All*	*M*	*F*	*18–24*	*25–34*	*35–44*	*45–59*	*60+*
I feel myself to be part of the new peace movement	6%	7%	6%	8%	12%	17%	6%	3%	3%
I sympathize with this new peace movement	46	59	58	61	70	60	59	59	54
I really have little sympathy for the new peace movement	28	20	20	20	10	15	18	24	24
I personally reject the new peace movement	19	13	16	10	7	7	14	12	19

SOURCE: Rückwirkungen weltpolitischer und weltwirtschaftlicher Ereignisse auf die politische Lage unseres Landes [The Political Situation of Our Country as Affected by Events in World Politics and International Economic Phenomena], SINUS, Munich, Oct. 1982.

West Germany—EMNID: The meetings and demonstrations against the deployment of the new missiles in our country are carried mainly by the peace movement. There are different views as to the people who give this peace movement its distinctive character and who are most influential. I have here a listing of a number of views that are frequently heard. Would you please tell me which of them you feel to be most correct.

	Deployment of Missiles								*Pro*	*Contra*
Nov. 24–Dec. 5, 1983	*All*	*M*	*F*	*–19*	*20–29*	*30–49*	*50–64*	*65+*	*(27%)*	*(57%)*
Brutal thugs who are mainly out for trouble and chaos	12%	12%	11%	9%	9%	9%	21%	9%	27%	3%
Worried citizens who consider the Federal Republic to be menaced by new nuclear missiles	31	27	34	43	32	27	23	43	13	45
Communists who wish to weaken the defense capabilities of the West	10	15	6	9	7	13	10	7	28	3
People who besides being opposed to new missiles wish to disarm generally	32	36	30	32	36	37	33	17	28	39
Have not yet heard anything about the peace movement	2	1	3	1	–	2	5	2	–	2
Don't know	10	7	11	5	10	7	6	22	3	6
No answer	3	2	4	1	5	5	2	1	1	2

West Germany—EMNID: Let us suppose it became known that within this year or next year U.S. troops would deploy mid-range missiles in this area and there were protests against the move. How would you react in such a case?

July 7–27, 1983, n = 1011	*All*	*M*	*F*	*–19*	*20–29*	*30–49*	*50–64*	*65+*
I would try to talk friends, co-workers, and neighbors out of protesting	15%	17%	13%	8%	16%	18%	13%	15%
I would do nothing for or against it	47	46	47	29	25	47	58	67
I would sign lists of protests	28	27	28	42	46	25	20	15
I would march in protest demonstrations	11	11	11	19	26	9	7	–
I would join a citizens' initiative against such a deployment	5	4	5	11	13	2	2	1
No answer	2	2	2	5	4	1	–	1

SOURCE: World Opinion Update, Nov. 1983.

West Germany—EMNID: What is your feeling: After this peace demonstration,* is the policy of the Federal Government likely to change greatly, somewhat, or not at all?

Oct. 15–Nov. 9, 1981, n = 1007	*All*	*M*	*F*	*–19*	*20–29*	*30–49*	*50–64*	*65+*
Greatly	6%	5%	6%	6%	7%	6%	6%	2%
Somewhat	35	33	37	44	45	34	27	32
Not at all	58	62	55	47	48	60	65	63
No answer	1	–	2	2	–	–	1	4

* Referring to the peace demonstration in Bonn, October 10, 1981.

Netherlands—Lagendijk (conducted for AVRO): On October 29 this year, a so-called peace demonstration will be held in The Hague. The demonstration is organized by the Labor Party, other leftist parties, D'66 [1966 Democratic Party], FNV [Federation of Netherlands Labor Unions], IKV [Interchurch Peace Council], the Stop the Neutron Bomb Group, and others. The slogans are: "No New Nuclear Weapons in Europe, Neither in the Netherlands Nor in Any Other Country," "Europe Free From Nuclear Weapons," and "Oust Nuclear Weapons From the World." The demonstration will aim specifically at preventing cruise missiles from being based in the Netherlands, irrespective of the outcome of negotiations. Are you favorably inclined toward this demonstration, or would you describe your attitude as skeptical, reserved? Do you definitely, possibly, or not intend to march for the peace demonstration?

		Sympathetic to Demonstration			
Oct. 7–10, 1983, n = 540	*Skeptical*	*Will Definitely March*	*May Possibly March*	*Will Not Join the March*	*No Opinion*
Total	38%	10%	10%	42%	–
Men	45	11	8	35	–
Women	31	10	11	48	1
18–34	29	15	17	40	–
35–49	44	10	9	37	–
50+	43	6	3	47	1

The Use of Nuclear Weapons

The Atlantic Institute for International Affairs—Louis Harris Poll

	Sep. 10–Nov. 4, 1983									*Apr. 1984*
	France	*Germany*	*Great Britain*	*Italy*	*Netherlands*	*Norway*	*Spain*	*USA*	*Japan*	*Greece*
The use of nuclear weapons is not acceptable under any circumstances, not even if we are attacked with nuclear weapons	27%	31%	24%	47%	42%	48%	61%	14%	58%	51%
Nuclear weapons should be used if we are attacked with nuclear weapons	52	42	61	28	36	45	24	66	18	40
If we are attacked with non-nuclear weapons, we would be justified in using nuclear weapons to end the war quickly	8	4	8	5	4	4	2	14	3	5
No answer/no opinion	13	23	7	20	18	3	13	6	21	4
n =	1023	943	1036	944	1003	1197	956	1246	1568	600

The Atlantic Institute for International Affairs—Louis Harris Poll: In the current debate on East-West nuclear weapons, which of the following best expresses your personal view about what the West should do:

	Sep. 10–Nov. 4, 1983									*Apr. 1984*
	France	*Germany*	*Great Britain*	*Italy*	*Netherlands*	*Norway*	*Spain*	*USA*	*Japan*	*Greece*
Give up all nuclear weapons regardless of whether the Soviet Union does	16%	23%	17%	35%	25%	15%	55%	4%	22%	38%
Introduce no more nuclear weapons, not even if the Soviet Union does	13	18	12	10	20	21	12	8	20	17

	Sep. 10–Nov. 4, 1983									Apr. 1984
	France	*Germany*	*Great Britain*	*Italy*	*Netherlands*	*Norway*	*Spain*	*USA*	*Japan*	*Greece*
Introduce just enough nuclear weapons to create a balance between East and West until an acceptable agreement can be concluded	47	39	62	30	38	55	16	63	21	37
Introduce more nuclear weapons than the Soviet union has introduced, in order to establish and maintain nuclear superiority	6	1	4	2	2	3	2	20	4	5
No answer/no opinion	18	19	5	23	15	6	15	5	33	3
n =	1023	943	1036	944	1003	1197	956	1246	1568	600

The Deployment of New Nuclear Weapons in Western Europe

USICA: Do you favor or oppose having new nuclear missiles that can reach the Soviet Union stationed in (Survey Country)?[a]

Oct. 1981	*Favor*	*Oppose*	*No Opinion*	
Great Britain	44%	45%	11%	*n* = 500
France	42	36	22	*n* = 480
West Germany[b]	28	44	27	*n* = 495
Italy	43	51	6	*n* = 531
Netherlands	34	57	9	*n* = 387
Norway	36	45	19	*n* = 283
Belgium[b]	14	84	1	*n* = 619

[a] For France and Norway ". . . stationed in Western Europe?"

[b] Percentages do not add up to 100% because of rounding.

SOURCE: Leo P. Crespi, "Trends in US Standing in West European Public Opinion," Feb. 1982.

Gallup International/Newsweek Poll: If the United States and the Soviet Union cannot agree on limiting nuclear weapons by the end of 1983, should NATO proceed with its plan to deploy Pershing II missiles and cruise missiles in Western Europe?

Jan. 1983	*Yes*	*No*	*Don't Know*	
United States	45%	31%	24%	*n* = 754
France	34	29	37	
West Germany	37	35	28	
Netherlands	33	51	16	
Great Britain	43	34	23	

SOURCE: Newsweek, Jan. 31, 1983. Copyright 1982 and 1983 by Newsweek Inc. All rights reserved. Reprinted by permission.

Do you think (Survey Country) should or should not allow the new American-controlled Cruise nuclear missiles to be based here?

Great Britain		*Should*	*Should Not*	*Don't Know*
Oct. 1981[a]	MORI	31%	59%	9%
Oct. 1982	SOC	31	58	11

Great Britain		*Should*	*Should Not*	*Don't Know*
Jan. 1983[a]	MORI	36	54	19
Jan. 28–Feb. 3, 1983	SOC	32	54	13
Sep. 1983	MORI	45	44	11
Oct. 1983,[a] $n = 1032$	MORI/The Sunday Times	43	51	6
Nov. 1983, $n = 999$	MORI/CND[b]	38	50	12
	Men	48	43	9
	Women	28	57	15
	18–34	35	53	12
	35–54	35	52	13
	55+	42	47	11
Nov. 9–15, 1983, $n = 960$	SOC/Daily Telegraph	41	48	11
Mar. 28–Apr. 2, 1984	SOC/CND	37	53	9
May 23–28, 1984	SOC/CND	35	56	9

[a] Question omitted "new American-controlled."

[b] CND = Campaign for Nuclear Disarmament.

West Germany—EMNID/Daily Telegraph: (Same question as above)

Oct.27–Nov. 3, 1983, $n = 937$	*Should*	*Should Not*	*Don't Know*
All	25%	48%	26%
Men	31	49	20
Women	20	48	33
18–34	15	55	30
35–44	24	52	23
45–64	32	46	22
65+	29	38	32

Great Britain—MORI/BBC TV: I am going to read out some statements and I would like you to tell me, from this card, how strongly you agree or disagree. Britain should not allow cruise missiles to be sited in this country.

May 6, 1983, n = 1090

Strongly Agree	*Tend to Agree*	*Neither*	*Tend to Disagree*	*Strongly Disagree*	*No Opinion*
24%	18%	10%	23%	21%	4%

West Germany—Demoskopie Allensbach: The so-called NATO dual resolution has now been in force for some time. Under this resolution, the NATO countries have agreed, on the one hand, to counterbalance the Soviet intermediate-range missiles by stationing equivalent missiles in Central Europe, and on the other hand to start negotiations with the Soviet Union on limiting nuclear weapons. In general, do you approve or disapprove of this dual resolution?

	Conducted for:	*Approve*	*Disapprove*	*Undecided*
May 1981	Capital	53%	20%	27%
Jan. 1982		52	22	25
Dec. 1982	German Federal Press Agency	51	25	24
Aug. 1983	FAZ	49	23	28

West Germany—EMNID: NATO has decided on the one hand to deploy mid-range U.S. missiles in Western Europe, beginning in 1983. On the other hand, NATO has made a move to the Soviet Union to start negotiations aiming at an agreement on disarmament, in which case NATO would partially or totally withdraw these missiles. What do you think of this decision?

July 7–27, 1983	*For*	*Indifferent*	*Against*	*Undecided*	*No Answer*
All	29%	11%	33%	24%	4%
Men	35	9	32	20	3
Women	23	13	33	27	4
–19	36	11	25	24	5
20–29	24	12	37	24	3
30–49	26	10	37	24	3
50–64	37	11	29	20	3
65+	21	16	30	28	5

SOURCE: World Opinion Update, Nov. 1983.

Netherlands—Lagendijk (Conducted for: AVRO): NATO some years ago took what is called the dual resolution. It was decided to modernize nuclear weapons and at the same time to negotiate with the Soviet Union on arms control. In case of a favorable development of these talks there would be no new nuclear cruise missiles. Pending the outcome of the negotiations, NATO has taken in hand preparations for deployment. Do you object or do you not object to preparatory acts which for the time being would be nonbinding, such as designating sites for stationing, and the like? Do you feel your place of residence to be or not to be a suitable area to be tentatively earmarked in this manner for the stationing of cruise missiles?

		Do Not Object		
Oct. 7–10, 1983, n = 540	*Object*	*Also in Own Place of Residence*	*But Not in Own Place of Residence*	*No Opinion*
All	48%	6%	39%	7%
Men	36	10	45	8
Women	58	3	33	6
18–34	49	7	37	8
35–49	46	8	39	7
50+	48	5	41	6

Netherlands—NIPO (conducted for Elseviers Magazine): If it were finally decided after due consideration to base new nuclear weapons in our territory within the context of NATO, next year,[a] would you consider this to be a good decision, a deplorable but acceptable decision, or an erroneous decision to which you are and continue to be opposed?

			Good	*Deplorable But Acceptable*	*Bad Remain Opposed*	*No Opinion*
1981	Jan. 26–27		17.5%	35.5%	39%	8%
	Mar. 8–9		18	32	40	10
	Mar. 16–17		18	36	37	9
	Apr. 21–22		15	35	42	8
	Apr. 27–28		16.5	39.5	38	6
	Aug. 31–Sep. 1		15	37	41	7
1982	July 12–13		12	38	45	5
	Sep. 20–21		11	42	41	6
1983	Jan. 17–18		13	41.5	37.5	8
	Mar. 7–8		11	46	35.5	7.5
	June 27–28	$n = 1054$	11	42	41	6
	July 11–12	$n = 866$	11	36	49	4
	Aug. 15–16	$n = 1207$	12	42	42	4
	Oct. 3–4	$n = 901$	12	45	39	4

[a] In 1983: "this year."

SOURCE: Elseviers Magazine, Dec. 4, 1982; Oct. 22, 1983.

If the government and the majority of Parliament should decide to station cruise missiles in our territory within the context of NATO, would you consider that a good decision? Or deplorable but acceptable? Or a wrong decision to which you are and continue to be opposed?

	Good	*Deplorable But Acceptable*	*Bad Remain Opposed*	*No Opinion*
Feb. 1984	16%	45.5%	35.5%	3%

SOURCE: Elseviers Magazine, Mar. 3, 1984.

Netherlands—Lagendijk (conducted for AVRO): Supposing the Geneva talks to be unsuccessful and the Soviet Union to keep its SS-20 missiles based, do you believe that Western Europe, including the Netherlands, should go ahead and station the new nuclear weapons, or do you believe this should not be done?

Oct. 7–10, 1983, n = 540	*Should Be Stationed*	*Must Not Happen*	*No Opinion*
All	42%	56%	2%
Men	54	43	2
Women	31	67	2
18–34	38	61	1
35–49	47	50	3
50 and over	43	54	3

Netherlands—NSS: The demonstration in The Hague has the major aim of preventing American missiles with nuclear warheads from being based in the Netherlands. No decision has as yet been taken on such deployment, but the government has indicated Woensdrecht as their site if it should be resolved to station them. With which decision would you personally agree: a decision indeed to station the cruise missiles in the Netherlands, or a decision not to station them?

	Should Be Stationed	*Should Not Be Stationed*	*No Opinion*
Sep. 5–6, 1983, n = 880			
Before peace demonstration	26%	56%	18%
Nov. 14–15, 1983, n = 950			
After peace demonstration	33	56	11

Netherlands—NSS (conducted for RVD): As you may know, the Government must still decide on whether American cruise missiles with nuclear warheads are to be stationed in the Netherlands. The Government has already designated Woensdrecht as their site of deployment if it should be resolved to station them. With which government decision could you personally associate yourself: a decision indeed to station the cruise missiles in the Netherlands, or a decision not to station them?

	Should Be Stationed	*Should Not Be Stationed*	*No Opinion*
Jan. 16–17, 1984, n = 954			
All	24%	63%	13%
Men	29	60	11
Women	19	66	15
18–24 years old	24	62	13
25–39	23	67	10
40–54	27	62	10
55 and over	23	60	17

Netherlands—Interview: As you probably know, the Netherlands government is shortly to decide on whether or not American cruise missiles will be stationed at the Woensdrecht base. Which decision do you believe the government ought to take: to refuse permission to station these missiles on Netherlands territory, or to grant permission to station these missiles on Netherlands territory?

	Grant Permission	*Refuse Permission*	*Don't Know*
May 1984, n = 750			
All	36%	52%	12%
Men	46	43	11
Women	27	60	13
–35 years	35	55	10
35–49	41	51	8
50+	34	49	17

Greece-Eurodim: If the SS-20's are not withdrawn, the installation of cruise missiles and Pershing-2's should or should not take place?

Oct. 1983, n = 600	
Should take place	18%
Should not take place	77
Don't know	5

Expected Developments on the Cruise Missiles in the Netherlands

Netherlands—NIPO/Elsevier: In your opinion, will this demonstration[a] be likely to yield any practical results in stopping nuclear missiles in the Netherlands, or do you think it is bound to be futile?

	Considered Futile	*Practical Result Expected*	*No Opinion*
Aug. 1983	72%	22%	6%
Sep. 1983	70	25	5
Oct. 1983	69	25	6

[a] Peace demonstration in The Hague, Oct. 29, 1983.

Netherlands—NSS (conducted for RVD): During the past few weeks there have been several demonstrations and other compaigns against the deployment of cruise missiles in the Netherlands. Which is your view: Would such actions still be capable of inducing the Government to reverse a decision to station the missiles, or do you believe they will have little or no effect?

	May Induce Gov't to Reverse Decision	*Little or No Effect*	*No Opinion*
Jan 16–17, 1984, n = 954			
All	12%	79%	8%
Men	13	82	6
Women	12	78	10
18–24	16	78	10
25–39	15	80	5
40–54	13	78	9
55+	8	80	12

Netherlands—Interact/Nieuwe Revu: We have been having a lot of commotion recently on the deployment in our country of so-called cruise missiles, which are missiles with a nuclear warhead. How would you estimate the odds of such missiles being stationed in the Netherlands?

	Great Likelihood	*Fairly Great Likelihood*	*Chances Are Against It*	*No Chance of Deployment*	*No Opinion*
Oct. 1982, all	13%	53%	28%	4%	2%
Men	15	54	24	7	–
Women	10	53	32	2	3

	Great Likelihood	*Fairly Great Likelihood*	*Chances Are Against It*	*No Chance of Deployment*	*No Opinion*
Oct. 1983, all	32	48	16	1	3
Men	40	43	12	2	3
Women	24	53	19	1	3

Netherlands—NIPO/Elsevier: What do you expect will happen: Is it going to be decided to station new nuclear weapons in the Netherlands within the context of NATO, or will this not come off?

	Aug. 1982	*Sep. 1982*[a]	*Jan. 1983*	*Mar. 1983*	*June 1983*	*July 1983*	*Aug. 1983*	*Oct. 1983*
Deployment expected	48%	70%	69%	72%	75%	82%	77%	80%
Will not come off	36	19	20	18	17	12	18	14
No opinion	16	11	11	10	8	6	5	6

[a] After publication of national election results.

Netherlands—NSS: How probable do you consider it all in all that the cruise missiles will eventually be stationed in the Netherlands?

	A Certainty; (Most) Probable	*Maybe Yes, Maybe No; No Opinion*	*(Most) Probably Not; Definitely Not*
Sep. 5–6, 1983, $n = 880$	67%	23%	10%
Nov. 14–15, 1983, $n = 950$	75	16	9
Jan. 16–17, 1984, $n = 954$ (conducted for RVD)	77	17	7
Men	83	11	6
Women	72	22	6
18–24	76	18	6
25–39	84	13	4
40–54	75	16	8
55+	73	19	8

Netherlands—Interact/Nieuwe Revu: Do you consider it right or wrong for vehement opponents of cruise missiles to attempt preventing their deployment, if need be by committing civil disobedience, such as the blockading of streets?

	Right	*Wrong*	*No Opinion*
Oct. 1982, all	35%	64%	1%
Men	26	74	–
Women	42	56	2
Oct. 1983, all	29%	71%	–
Men	28	72	–
Women	30	70	–

Netherlands—Lagendijk (conducted for AVRO): Let us suppose that Parliament, the highest quorum in our country, were to agree on the deployment of a given number of cruise missiles in the Netherlands. It is to be expected that there will be no large majority in Parliament in favor of such deployment. On the basis of our democratic rules it would nevertheless be a legally valid decision. Will you accept a decision of Parliament to station nuclear missiles? Does your nonacceptance of such a decision of Parliament mean that you will personally take part in campaigns, or is that not your conclusion? If you join the protesters, will you take part exclusively in legal actions, such as political pressure, strikes, demonstrations, and the like? Or are you also prepared to engage in hard forms of resistance, such as blockades, sabotage, rioting?

		Do Not Accept Decision of Parliament			
Oct. 7–10, 1983 *n = 540*	*Accept Decision of Parliament*	*Will Not Actively Resist*	*Will Resist, but Only by Legal Means*	*Will Resist, Also in Hard Campaigns*	*No Opinion*
Netherlands public	54%	25%	17%	1%	3%
Men	62	21	13	2	2
Women	47	29	20	1	4
18–34	50	21	23	3	3
35–49	56	26	16	1	1
50+	57	29	10	–	4

Netherlands—NIPO (conducted for Elsevier): Do you feel that protesters and campaigners should attempt, if necessary by force, to keep the cruise missiles away from Woensdrecht?

	Keep Away by Force	*Refrain from Action*	*No Opinion*
Aug. 15–16, 1983 *n* = 1207	12%	83%	5%

SOURCE: Elseviers Magazine, Oct. 22, 1983.

Netherlands—NSS (conducted for RVD): Whether the cruise missiles will be based in the Netherlands depends ultimately on the concurrence by Parliament with a decision of Government to deploy them. If a majority of Parliament should approve the decision of stationing the missiles, with a large part of the population being opposed to their being stationed, would you acquiesce in it, or would you not accept it passively and, for example, take part in campaigns?

	Would Acquiesce	*Would Not Accept Passively, Would Take Part in Campaigns*	*Don't Know*
Jan. 16–17, 1984			
Total	51%	43%	5%
Men	49	45	5
Women	53	41	6
18–24	56	35	9
25–39	47	49	5
40–54	57	37	6
55+	50	46	4

NOTE

ONLY THE FOLLOWING PAGES OF THIS

PUBLICATION HAVE BEEN FILMED

Cov. 1, p. 141-252, 264-274.

A0610-1

POPOAE 49(2) 143–284 (1985)
ISSN 0033–362X

Public Opinion Quarterly

WOMEN'S CHANGING ROLES
Women's Gender Consciousness — *Patricia Gurin*

Myth and Reality: The ERA and the Gender Gap in the 1980 Election — *Jane J. Mansbridge*

Gender Differences in Voting for Female Candidates: Evidence from the 1982 Election — *John F. Zipp and Eric Plutzer*

Like Mother, Like Daughter: Intergenerational Transmission of DK Response Rates — *Ronald B. Rapoport*

Expectation/Disillusion and Presidential Popularity: The Reagan Experience — *Lee Sigelman and Kathleen Knight*

Racism, Rational Choice, and White Opposition to Racial Change: A Case Study of Busing — *McKee J. McClendon*

Face-to-Face Interviews and Mailed Questionnaires: The Net Difference in Response Rate — *John Goyder*

The Instability of Response Sets — *C. Harry Hui and Harry C. Triandis*

THE POLLS: America's Most Important Problems, Part I: National and International — *Tom W. Smith*

COMMENTS AND LETTERS · BOOK REVIEWS
BOOK NOTES

SUMMER 1985

The Public Opinion Quarterly

VOLUME 49 SUMMER 1985 NUMBER 2

Women's Gender Consciousness *Patricia Gurin* 143

Myth and Reality: The ERA and the Gender Gap in the 1980 Election *Jane J. Mansbridge* 164

Gender Differences in Voting for Female Candidates: Evidence from the 1982 Election *John F. Zipp and Eric Plutzer* 179

Like Mother, Like Daughter: Intergenerational Transmission of DK Response Rates *Ronald B. Rapoport* 198

Expectation/Disillusion and Presidential Popularity: The Reagan Experience *Lee Sigelman and Kathleen Knight* 209

Racism, Rational Choice, and White Opposition to Racial Change: A Case Study of Busing *McKee J. McClendon* 214

Face-to-Face Interviews and Mailed Questionnaires: The Net Difference in Response Rate *John Goyder* 234

The Instability of Response Sets *C. Harry Hui and Harry C. Triandis* 253

COMMENTS AND LETTERS

On Preelection Polling *Nick Panagakis* 261

Response to Panagakis *Richard Day* 262

Errata 263

THE POLLS: America's Most Important Problems, Part I: National and International *Tom W. Smith* 264

BOOK REVIEWS *Marc B. Glassman, Editor* 275

Peter H. Rossi, James D. Wright, and Andy B. Anderson, *Handbook of Survey Research* *Charles D. Cowan* 275

Ithiel deSola Pool, *Technologies of Freedom* *Jack Lyle* 276

David Butler and Dennis Kavanagh, *The British General Election of 1983* *Robert M. Worcester* 279

BOOK NOTES *Phyllis M. Endreny, Editor* 282

NOTICE TO CONTRIBUTORS

Three copies of each article should be submitted. Authors are urged to submit original typescripts. Copies of manuscripts which are reviewed become the property of the journal, and unused copies will be returned only if the author provides a stamped self-addressed envelope. Bibliographic-references, footnotes, and tables should be typed to conform to the printed style of this journal.

All submissions that make use of survey data not archieved at the Inter-University consortium for Political and Social Research or the Roper Center must contain the following information: (1) a definition of the population from which the sample was drawn and a statement of the selection procedure sufficient to permit at least approximate replication. (2) The response rate and details of its calculation. (If response rate is not appropriate, the number of refusals.) (3) Dates the survey was conducted. (4) Exact wording of all questions used in the analysis.

Upon acceptance of an article by the journal, the author(s) will be asked to transfer copyright in the article to the publisher. This transfer will insure the widest possible dissemination of information under the U.S. copyright Law.

The PUBLIC OPINION QUARTERLY is hospitable to all points of view, provided that the material presented will help to illuminate problems of communication and public opinion. The American Association for Public Opinion Research assumes responsibility only for material signed by an officer of the Association in his official capacity. Statements of fact and opinion are made on the responsibility of the authors alone and do not represent the views of the editors.

The PUBLIC OPINION QUARTERLY has been registered with the Copyright Clearance Center, Inc. consent is given for copying of articles for personal or internal use, or for the personal or internal use of specific clients. This consent is given on the condition that the copier pay through the Center the per-copy fee stated in the code on the first page of each article for copying beyond that permitted by the U.S. Copyright Law. If no code appears on an article, the author has not given broad consent to copy and permission to copy must be obtained directly from the author. All articles published prior to 1978 may be copied for a per-copy fee of $1.00, also payable through the Center. The consent does not extend to other kinds of copying, such as for general distribution, resale, advertising and promotional purposes, or for creating new collective works.

Women's Gender Consciousness

PATRICIA GURIN

FAMILY life and the roles of women and men within and outside the family have changed impressively over the past 30 years. Large numbers of women, even those with young children, have moved into the labor force. By 1979 women comprised 51 percent of the students in higher education institutions, up from 38 percent in 1963.[1] And while the division of labor within the family is far from equal, some changes toward equality have occurred.[2] Attitudes toward gender roles are now more

---

[1] Table 80, Digest of Educational Statistics, 1981, Washington, D.C.: National Center for Education Statistics, 1981. Figures for 1980 show that the female edge had risen to 52 percent, although women have not reached parity among students receiving bachelor's or doctoral degrees (Scientific Manpower Commission, 1983).

[2] The best data on this are from time-budgets, but these have been kept for research purposes for only a short while. Self-reports about responsibility for household chores show that household division of labor has changed most among the young, better-educated husbands and wives, and families with young children (Goode, 1982).

Abstract Four dimensions of group consciousness that figure prominently in relative deprivation and resource mobilization/solidarity theories of social movements were measured in three national cross-section sample surveys over the decade of the 1970s by the Institute for Social Research. These dimensions—identification, discontent, withdrawal of legitimacy, and collective orientation—were applied to the gender consciousness of women. In all but one of these dimensions, women's gender consciousness was comparatively weak. It was not as pronounced as the group consciousness of other subordinate categories, nor was it distinctly subordinate because men expressed similar views. A structural interpretation of this comparative weakness is offered. Women did become more group conscious between 1972 and 1983. The increase was particularly pronounced with regard to their discontent about the relative political power of men and women and their views about the legitimacy of gender disparities.

Patricia Gurin is Professor of Psychology and Faculty Associate at the Institute for Social Research, The University of Michigan, and a Resident Scholar at the Russell Sage Foundation, New York. The author acknowledges the helpful criticisms of an earlier draft by Arland Thornton and Mary Corcoran of The University of Michigan, Robert K. Merton of Columbia University, Faye J. Crosby of Yale University, and Rhea Kish of Ann Arbor, Michigan.

Public Opinion Quarterly Vol. 49:143–163 © by the Trustees of Columbia University
Published by Elsevier Science Publishing Co., Inc. 0033-362X/85/0049-143/$2.50

egalitarian; for example, they are more favorable toward women working outside the home, combining jobs with motherhood, and otherwise participating in the public arena. In a panel study of women, Thornton and Freedman (1979) found shifts ranging from 16 to 35 percent between 1962 and 1977 toward greater support for nontraditional roles and division of labor. A follow-up of these women in 1980 showed these general trends continuing despite a more conservative political environment (Thornton, et al. 1983). Data from the last 30 years of election studies at the Institute for Social Research also show that the male edge in turnout has virtually disappeared (Baxter and Lansing, 1983) and that women are now more active in electoral campaigns than they were in the 1950s (Andersen, 1975; Welch, 1977).

To what extent have these status and attitude changes been accompanied by more general ideological awareness of how gender shapes the lives and fortunes of men and women? Do more women now recognize that they are objectively deprived economically and politically? Do more women believe that women are treated categorically? Have women become more aggrieved about their relative influence in the public arena? Have more begun to question the legitimacy of gender disparities, such as their continuing wage gap?

There is reason to suspect little change, despite the seemingly dramatic changes selectively depicted by the media. For one thing, the American scene does not promote group-conscious beliefs among deprived, subordinated social categories. During most of its history the United States has been characterized by an expanding economy, large geographical population shifts, and enough intergenerational mobility to allow Americans, aided by an ideology of meritocracy, to presume more status changes than have actually existed. These conditions all encourage individualism, not solidarity or the recognition of group disparities and awareness of structural forces that impede achievements for some social categories. Instead, Americans tend to morselize experience (Lane, 1962), fail to connect their own experiences to those of others of like status (Crosby, 1982), and firmly believe in a just world where individuals largely receive what they deserve (Lerner, 1981). Thus, the very idea of categorical treatment flies in the face of widely held values.

Current theory can help us understand why women have not developed group consciousness as readily as other subordinate categories in America. Women lack many of the structural conditions that promote bonds to others in the same subordinate category (Merton, 1968; Williams, 1975). In the first place, gender inequality, while present and doggedly persistent in some areas, is not as marked as racial inequality, and perception of inequality is more difficult when it is not extreme. Further, it is lower in the United States than in most other Western industrialized nations. Data from the United States, Finland, Switzerland, The Netherlands, Great Britain, Italy, Germany, and Austria show the United States and Finland with the smallest gender gap in years of schooling and personal income, while the gap is especially pronounced in Germany and Austria (Jennings and Farah, 1980). Second, the economic fates and the claims to prestige of men and women are inextricably intertwined. Both sexes jointly share their gains and losses as members of ethnic groups or economic classes. Third, they agree on many fundamental assumptions about life and society, not only because of their common fates as husbands and wives, but because of their early socialization as brothers and sisters. This does not negate the existence of gender differences in socialization. Both systematic research and personal accounts are replete with examples of the differing treatment of boys and girls in the family and in school. Still, both are given many of the same messages about basic societal values, including individualism and a theory of justice in which the distribution of rewards is to be expected from the distribution of individual talents (Brickman, 1980). Cleavage and conflict rarely develop between groups that share such fundamental values (Coser, 1956).

Fourth, and probably the most profound inhibitor of gender consciousness among women, is the structure of their relations with other women and with men. Solidarity and recognition of group deprivation are fostered when category members interact most frequently with each other and only moderately with the outgroup; when intimate interaction is confined to those within the ingroup; when relations with the outgroup are predominantly competitive or, at the very least, not intimate (Williams, 1975). Women's relations with each other and with men could not be more different from this pattern. Their relations with each other, while frequent and intimate, are at times competitive. At the same time, nearly all women have, or have had, close emotional and familial ties to men. Further, some women derive their social status and prestige entirely from their roles as wives and homemakers, which provide protections, meet dependency needs, and even allow flexibility and control when traditional arrangements work out ideally. These traditional roles are not without rewards for women. As a consequence some women have a vested interest in the social unit that also strengthens their unequal status (Goode, 1982). Men, individually and collectively, also have had a stake in justifying and maintaining the inequality in the structure of male-female relations. Until recently they have controlled the legitimating mechanisms that justify male dominance, including the production of theories about gender differences (Epstein, 1984). Moreover, for many women, associations outside the family involve cross-cutting ties, in part with men and in part with women of more priviledged statuses, which mute their recognition of the deprived circumstances of women as a class.

These structural conditions in their own lives obstruct strong gender consciousness and ideological change among women, despite shifts in status and in general attitudes toward gender roles. For these reasons, this paper examines gender consciousness in women, including trends over the 1970s into the more conservative years of the early 1980s.[3] Comparisons are drawn with the race consciousness of blacks, the age consciousness of older people, and the class consciousness of people in working-class jobs.

Data come from three national cross-section sample surveys carried out by the Institute for Social Research at The University of Michigan. Data from males in these surveys are also analyzed.

Components of Group Consciousness

The concept of group consciousness is complex, but four main components are suggested by two current theories of collective action: the relative deprivation and the solidarity/resource mobilization theories.[4] Invoking different terms, each seeks to explain how and why change-oriented action emerges in a constituency.

The first component—*collective orientation*—assumes that the group desires change in rank or power because either it has been subordinated or its dominance has been challenged. We are interested here in comparing women's attitudes toward collective action with that of other subordinated groups.[5] In our definition of group consciouness, collective action is geared to change, and in this sense gender consciousness is feminist.

---

[3] Colleagues Arthur Miller, Gerald Gurin, Aloen Townsend, and Oksana Malanchuk collaborated in developing the concepts and measures, first described in Gurin, et al., 1980.

[4] In reacting against the psychological emphases in relative deprivation theories, many resource mobilization theorists initially discounted individual factors altogether. They looked instead to organizational resources, social control policies of authorities, configuration of allies and enemies, and the structure of the political economy to explain the emergence of collective action (Oberschall, 1978). Lately, some theorists within this tradition have reinstituted social psychological variables, in particular shared identity, sharpness of perceived group boundaries, and extensiveness of social networks within a constituency (Fireman and Gamson, 1979; Tilly, 1978).

[5] Group consciousness, of course, holds in dominant groups as well. It justifies advantage, gives legitimacy to their social and economic superiority, and evokes action aimed to perpetuate that advantage. Conscious articulation of a dominant group ideology occurs primarily when it is challenged by deprived groups. When not under challenge, dominant groups have less need of a systematic theory about intergroup relations because the status quo is presented as a natural state of affairs, good not just for the dominant group but for society as a whole (Mannheim, 1972).

The other three components are associated with individuals' feelings and perceptions about their membership in a social category. The second component—individuals' *discontent* about the power of their social category—reflects the central concern of relative deprivation theories. Group consciousness and the emergence of collective action relate to the intercategory or intergroup connections that people make with those of dissimilar status or between those in their category and an outgroup (Williams, 1975). The critical question is what structural and psychological forces encourage individuals to draw these comparisons so as to become collectively discontented.

One structural condition that encourages or blocks outward comparisons is the degree of legitimacy ascribed to the prevailing social structure (Merton and Kitt [Rossi], 1950). This suggests a third component: how category members appraise the *legitimacy of disparities*. Disparity will not be experienced as deprivation unless members believe it springs from illegitimate forces. Group disparities may be considered fair and just if they are believed to develop legitimately from the inadequacies of individual category members. The same disparities will be judged unjust if they are believed to result from structural barriers that hold back category members.

The fourth component—*identification*—reflects a recognition of shared values and interests that turns a category into a collectivity (Merton, 1968). Most relative deprivation theories do not explicitly address how the sense of collectivity strengthens recognition of shared deprivation or provides a political interpretation for deprivation.[6] Solidarity theories, however, stress the motivational importance of members' awareness of common feelings, values, and interests. When members share a strong identity and have strong bonds to others in the category, recruitment, mobilization, and organization are more easily achieved (Oberschall, 1978; Tilly, 1978; Fireman and Gamson, 1979). In addition, identification has cognitive effects that are important for political consciousness.[7] By sharpening the salience of groups and the likelihood of using an intergroup perspective for comparison, identification heightens recognition of collective deprivation and encourages withdrawal of legitimacy because deprivation, when perceived as a common rather than individual condition, is more easily interpreted in structural and political terms.

---

[6] Williams (1975) is a clear exception in that he credits solidarity with explaining the experience of inequality as deprivation rather than as the nature of things.

[7] Solidarity theories have had little to say about the cognitive consequences of identification. In the sense that awareness of group boundaries is considered an aspect of solidarity itself, they do deal with intergroup perception. However, they pay little attention to the ways in which solidarity, once present, influences perceptions of the membership and nonmembership groups. Overt conflict is credited with sharpening boundaries and exaggerating perception of difference (Coser, 1956). The idea that solidarity itself could result in sharpening the perception of boundaries, even without conflict, is suggested, however, by current research on social cognition. (See especially Tajfel, 1978, and Stephan, in press.)

Procedures

SAMPLES

Three national probability samples of adults were surveyed by the Institute for Social Research. The 1972 national election study (NES) included 1537 women and 1168 men, the 1976 NES, 1662 women and 1208 men, all interviewed face-to-face; two waves of the monthly consumer surveys in 1983 included 730 women and 588 men interviewed by telephone. In addition, one panel of 742 women who participated in both election surveys is used for analyzing which subgroups changed most and least.

MEASURES

Collective orientation was measured from responses to forced-choice questions. These probed views about action strategies women should follow to improve their market statuses. One alternative advocated that women work together to change the laws and social processes that cause job and wage inequality; another advocated individual effort by each woman to get ahead on her own. These questions assume that gender disparities in the labor market should be reduced. As noted previously, our model of group consciousness does not attempt to account for collective strategies that resist change in existing gender roles and relations. Desirability of change is also assumed in a third question about collective orientation. This measures attitude toward the "Women's Liberation Movement," using a feeling thermometer on which 0 to 49 degrees represented cold or unfavorable feeling, 50 a neutral feeling, and 51 to 100 warm or favorable feeling. These measures were collected in 1972 and 1976 but not in 1983.

Discontent was measured from evaluations of the influence of various groups in American society. The interviewer stated that "some people think that certain groups have too much influence in American life and politics, while other people feel that certain groups don't have as much influence as they deserve. On this card are three statements (too much influence, just about the right amount, too little influence) about how much influence a group might have." Absolute discontent, represented by the statement that "women" have "too little influence" was assessed in all three surveys. Relative discontent, the discrepancy between evaluations of the power of "women" and the power of "men," was assessed beginning in 1976.

One measure of *withdrawal of legitimacy* was derived from responses in 1972 and 1976 to three forced-choice questions in which respondents were asked to identify the causes of gender differences in income, occupational status, and general position in American life. For each question, one alternative attributed these differentials to systemic obstacles and institutional arrangements such as market discrimination or poor schooling; the other to women's personal deficiencies such as lack of ambition or preferences and natural endowments antithetical to economic success. Since forced-choice questions are difficult when interviewing by telephone, the 1983 interviews repeated Likert-type versions of these attributions that had also been asked in 1972. We also included a measure of attitude toward gender role differentiation,which factor analyses confirmed was part of the legitimacy issue. Role attitude was measured in all three surveys by asking respondents to place themselves on a seven-point scale, where one represented the opinion that "women's place is in the home" and seven that "men and women should have an equal role in running business, industry, and government."

Gender identification/sense of collectivity was measured by asking respondents to look at a set of 16 category labels that included "women." The interviewer asked, "Which of these groups do you feel particularly close to—people who are most like you in their ideas and interests and feelings about things?" Respondents were then asked to look at the list again and choose the *one* to which they felt closest. Degree of identification was operationalized as: (1) *not identified*, women who did not pick the label "women" as a category to which they felt even close; (2) *identified*, women who said they felt close but not closest to "women"; (3) *closely identified*, women who said they felt closest to "women." This measure was not used in the 1983 telephone survey.

Factor analyses of the measures that were phrased identically in 1972 and 1976 confirmed our expectation of finding the four dimensions defined above. We present analyses of both the single items and unit-weighted indexes formed on the basis of these factor analyses.[8]

---

[8] These analyses were carried out using LISREL IV and are described in detail elsewhere (Gurin, 1984). Bentler and Bonett's index of incremental fit, Delta, was used to evaluate the measurement model. A model is generally considered acceptable when Delta is greater than .90. Delta for the measurement model was .93 in 1972 and .96 in 1976. Dawes and Corrigan (1974) show that unit-weighted measures and factor scores in which items are weighted by their factor loadings are very highly correlated, and they recommend the former as being more easily interpreted.

Strength of Gender Consciousness

COMPARISONS AMONG SUBORDINATE CATEGORIES

Questions similar to those representing the four components of gender consciousness were asked of people over 60 about "older people," of people in working class occupations about "workingmen,"[9] and of blacks about "blacks," to compare the strength of women's gender consciousness to the group consciousness of other subordinate categories.

Many of the structural conditions that constrain gender consciousness among women do not operate as strongly in other subordinate categories, especially among blacks. The amount of objective racial inequality and the reality of residential segregation help blacks recognize disparity. Relationships with the outgroup are more typically emotionally distant; and the black community has been organized for directed social change for a longer time. Prior organization and leadership are critical resources for developing and mobilizing group consciousness for collective action. Political scientists credit solidarity and organization as the resources that best explain political participation rates among blacks that exceed expectations based merely on their educational and economic characteristics (Verba and Nie, 1972; Miller, et al., 1981).

Thus it is not surprising that women were not as group-conscious as blacks in either 1972 or 1976. But it is surprising that most components of women's gender consciousness were weak compared with the class consciousness of people in blue-collar jobs and even the age consciousness of people over 60. Since the comparative results for 1972 are presented elsewhere (Gurin, et al., 1980), and the 1976 results are essentially the same, we present these comparisons in summary rather than tabular form

Women's *identification* with other women was much less widespread. Compared to women, three times as many blacks and people 60 or older, and two-and-a-half times as many people with blue-collar jobs "closely identified" with their own groups. The edge of blacks relative to women was approximately as large in 1976, while the edge of older people and blue-collar workers was somewhat less than in 1972.

Identification can also be examined by looking at those women who closely identify as women when they objectively belong to at least one other subordinate category. Here the concern is with competing choices among subordinate memberships. Compared to such women in 1972, seven times as many blacks who belonged to at least one other subordinate category identified closely with "blacks." Five times as many older people and three times as many blue-collar workers did so. The comparisons in 1976 were approximately the same, all showing the more exclusive and stronger identification in the other subordinate categories. Thus for all comparisons, identification was less widespread among women.

---

[9] Use of the term "workingmen" instead of "working class people" is an unfortunate example of our lack of feminist consciousness when these questions were first asked in the 1972 national election study. The term was carried along to the later surveys to assure being able to assess response shifts across time.

Discontent and *collective orientations* were also weaker among women. For these dimensions, comparisons were made between members who closely identified with their subordinate categories. This provides comparisons of independent samples because respondents could choose only one category that they feel "closest to."

Compared to closely identified women in 1972, one-and-two-thirds as many closely identified older people and blacks were discontented with their category's power and influence. Discontent among closely identified blue-collar workers was not very strong and was approximately the same as among women. Closely identified members of the other subordinate categories were also stronger than women in their collective commitments. On a seven-point scale probing what members of their category should do "to have influence and get the things they want," the choices ranged from "organize as a group" to "work as individuals." Compared with closely identified women in 1972, two-and-a-half times as many blacks, nearly twice as many blue-collar workers, and one-and-a-half times as many older people selected the two most extreme collective positions on the scale. Women continued to show nearly the same relative weakness in the 1976 comparisons.

Legitimacy was the one aspect of gender consciousness about which closely identified women were politically conscious. They were as aware as closely identified blacks of structural causes of their own group's disadvantage in the labor market. (The legitimacy measure compared here does not include the item measuring women's attitudes toward gender roles, since blacks were not asked a comparable question.) In both 1972 and 1976, women's average scores on an index measuring causal attributions for gender disparities were nearly identical to those of closely identified blacks.

MALE-FEMALE CONSENSUS

In both Europe and the United States men and women share similar perspectives about the role of women in society and specifically in politics. Jennings and Farah (1980) have studied this in the United States and seven Western European countries with data collected in 1974. The results are clear. Women were objectively deprived everywhere. They had lower personal incomes, somewhat less education, and less employment experience than men. On the basis of some theories, such relative deprivation should lead to discontent. But nowhere were women much aggrieved about their place in politics, nor any more positive than men about the women's movement, except in Finland. Everywhere they shared much the same views about political agendas and the same partisan orientations. Data collected in 1983 in the 10 countries of the European Community show there is still remarkable consensus in the opinions of men and women on the roles the two sexes should play in the economy, politics, and the family (Commission of the European Communities, 1984). American polls taken since the early 1960s have always shown that men are somewhat more supportive than women of role equality (Shapiro, et al., 1984; Goode, 1982). Historically, American men and women have also held similar party and candidate preferences. The gender gap in the 1980 presidential election, in which fewer women than men voted for Reagan, was hailed because it was the first time that even a gap as large as 10 percent had appeared (Baxter and Lansing, 1983). And even in 1980, men and women differed as much as 10 percent on only 4 of the 14 policy issues probed in Michigan's national election study (Lake, 1982). Gender differences in political attitudes did increase during the first four years of the Reagan presidency (Frankovic, 1982; Blydenburgh and Sigel, 1983), although not in attitudes specifically toward role equality.[10]

Crosby's (1982) study of relative deprivation likewise shows consensus between women and men in evaluating the employment situations of women in the United States. Both agreed on the extent to which working women get what they want and deserve from their jobs, fare better or less well economically than 10 years earlier, will experience gains in the next five years, and are held back by their own personal deficiencies from obtaining good jobs, advancing on the job, and getting the pay they deserve. They differed only in that men less often perceived a gender gap in wages and benefits. Otherwise, their perceptions were remarkably similar. The critical differentiator was not gender but *gender role*. Employed women were consistently the most aware of women's collective deprivation in the labor market. Homemakers were the least critical, even less than men, of women's relative job deprivation.

---

[10] The controversy over whether public opinion polls usually reveal gender similarity or differentiation hinges largely on the definition of what is a meaningful difference. Shapiro and colleagues (1984) review responses to 267 policy questions, half of which were repeated three or more times, in polls carried out between 1964 and 1983 by several survey organizations. Confirming other studies (Smith, 1984; Pomper, 1975), they uncovered the largest gender differences (on average 8 percent) on issues that involved force or violence. These differences were also highly reliable, found in two-thirds of the polls. On five other issue categories, gender differences ranged from 2 to 6 percent. Since differences of 6 percent usually were statistically significant and were found in one-third of the polls, Shapiro and colleagues reject the common view that gender differences are small and inconsequential. The critical question is what size difference strikes the analyst as consequential, either compared to other bases of possible cleavage in society or because the difference affects the vote or other behaviors.

The lack of gender differentiation in analyzing women's position in society is also remarkable in our data. If the gender consciousness measures are taken as valid only when men and women differed on them, only the discontent measures would fit that criterion. However, the other measures prove to be valid also, since they are theoretically derived and have construct validity, as seen in the comparisons given earlier for the four subordinate categories. Moreover, they capture enormous differences between blacks and whites. Where there is a genuine cleavage in society, these measures tap it.

The 1976 data in Figure 1 show impressive racial differentiation and impressive gender consensus. (The 1972 data cannot be used because the survey did not ask about the power of whites and men.) Differences between men and women and between blacks and whites are shown for discontent, legitimacy, and collective orientations. For discontent, the

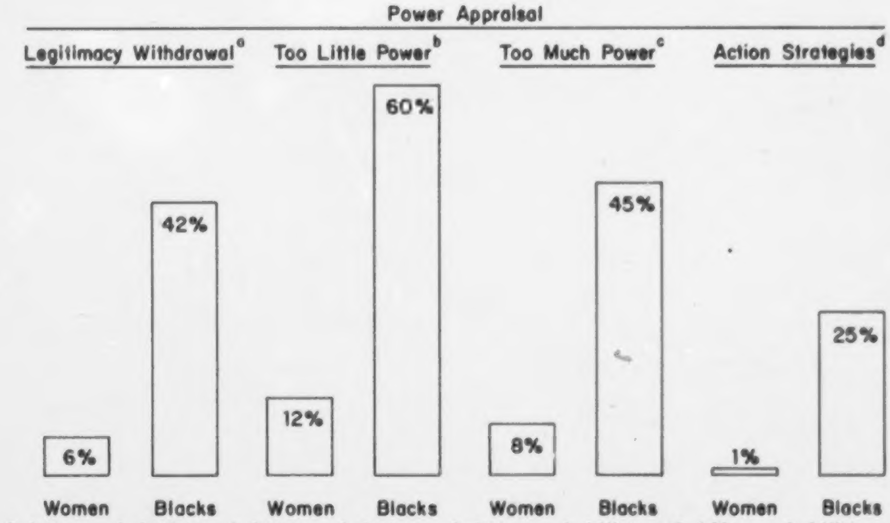


Figure 1. Similarity in Men's and Women's Views, Compared to Differentiation in Whites' and Blacks' Views, 1976

[a] % Difference between subordinate and superordinate in endorsing three-quarters of the alternatives blaming the system.
[b] % Difference saying subordinate has "too little" power.
[c] % Difference saying superordinate has "too much" power.
[d] % Difference endorsing both collective strategies.

data compare men's and women's evaluations of the power each ought to have in life and politics; blacks and whites made the same kind of evaluations for each of their groups. For legitimacy, causes men and women attributed to gender disparities were compared; similar comparisons were made for blacks and whites of their explanations for racial disparities in wages and occupational status. Although the phrasing of the items on the gender and racial indexes is not identical, each had the same number of forced-choice items with one alternative blaming structural forces and the other personal deficiencies. The same is true for collective orientation indexes, with one alternative advocating collective action by women (or by blacks), and the other favoring individual action.

The only statistically significant difference between men and women pertained to evaluations of power. And that lone difference is dwarfed by the extremely large differences between blacks and whites. Sixty percent more blacks than whites said blacks have "too little" power, and 45 percent more blacks said whites have "too much" power. Blacks and whites also had highly discrepant views about the causes of race disparities and blacks' strategies for reducing them. In contrast, men and women held nearly identical opinions about the individual and structural forces that produce gender disparities and strategies for women to improve their statuses.

Trends, 1972–1983

AGGREGATE SHIFTS AMONG WOMEN

All of the single items and indexes that were measured identically for any two time periods are presented in Table 1. In many respects, women's gender consciousness became stronger during the years following the 1972 survey. The largest shifts occurred in women's appraisals of the power men ought to have and of the legitimacy of labor market disparities.

We look first at identification. Change is clarified by looking closely at the proportions shifting out of specific categories of the identification measure. The major change was movement from nonidentification to moderate identification (choosing women as a category they felt "close" to). However, this movement proved not unique to gender. Women generally tended to identify with more social categories in 1976 (on average one more). Thus, movement toward feeling "close" to women says little that is particular to gender identification. The number who chose women as the "closest" category is a more compelling indicator because interviewees could select only *one* category as the closest identification. This indicator showed stability, not growth, in identification.

Approval of collective strategies was both weak and stable in this

Table 1. Women's Endorsements of Items Measuring 4 Components of Gender Consciousness

	Percent Endorsement		
	1972	*1976*	*1983*
IDENTIFICATION			
Chose women as the category to which felt:			
"Closest"	10%	12%	—
"Close"	36	48	—
Not identified	54	40	—
	100%	100%	
COLLECTIVE ORIENTATION			
Felt "warm" toward Women's Liberation Movement (thermometer rating of 51°–100°)	36%	46%	—
Chose: "It is not enough for a woman to be successful herself; women must work together to change laws and customs that are unfair to all women."	45	50	—
Rather than: "Women can best overcome discrimination by pursuing their individual career goals in as feminine a way as possible."			
Chose: "Only if women organize and work together can anything really be done about discrimination."	19	20	—
Rather than: "The best way to handle problems of discrimination is for each woman to make sure she gets the best training possible for what she wants to do."			
Summary index: Scores of 4 or 5 on 5-point index, 5 = high collective orientation	20	21	—
WITHDRAWAL OF LEGITIMACY			
Attitude toward equal roles: Endorsed 3 most extreme positions on 7-point scale advocating equal roles for men and women	47	51	53
Forced-choice causal attributions about disparity			
Chose: "Many qualified women can't get good jobs; men with the same skills have much less trouble."	61	65	—
Rather than: "In general, men are more qualified than women for jobs that have great responsibility."			
Chose: "Our society, not nature, teaches women to prefer homemaking to work outside the home."	44	61	—
Rather than: "By nature women are happiest when they are making a home and caring for children."			
Chose: "Men have more of the top jobs because our society discriminates against women."	48	62	—
Rather than: "Men have more of the top jobs because they are born with more drive to be ambitious and successful than women."			
Summary index: Scores of 4 or 5 on 5-point index, 5 = high collective orientation	43	56	—
Likert-type measures attributing disparity to the greater natural endowments of men			
Strongly *disagreed* that:			
"Men are born with more drive to be ambitious and successful than women."	31	—	37
"In general, men are more qualified than women for jobs that have great responsibility."	24	—	51
"By nature women are happiest when they are making a home and caring for children."	12	—	26
Summary index: Scores of 4 or 5 on 5-point index, 5 = reject "natural endowments" explanation	36	—	55
Likert-type measures attributing disparity to structural forces			
Strongly *agreed* that:			
"Many qualified women can't get good jobs; men with the same skills have much less trouble."	41	—	52
"Women have less opportunity than men to get the education for top jobs."	17	—	13
"Our schools teach women to want the less important jobs."	12	—	12
Summary index: Scores of 4 or 5 on 5-point index, 5 = acceptance of structural explanations	35	—	35
DISCONTENT			
Said:			
Women have "too little power"	35	40	43
Men have "too much power"	—	25	46
Women have "too little" relative to men	—	45	57

period, although women did increasingly favor the Women's Liberation Movement. Somewhat more women felt "warm" toward the movement in 1976 than in 1972. However, this increase was accounted for by the larger numbers of women who gave the movement a lukewarm rating (between 51 and 75 degrees). The proportion who rated it more warmly than 75 degrees did not increase. Depending on the item, since only 20 to 50 percent preferred collective to individual action even when they had become somewhat more approving of the women's movement, it may be that their increased warmth reflected approval of its goals, not its means.

Other aspects of gender consciousness shifted more on average. Women especially changed their views of the causes of disparities. On the forced-choice measures, 4–17 percent more women in 1976 expressed a structural view of disparity. This greater acknowledgment that the system discriminates against women is quite impressive since over the same period women became *less* convinced that discrimination is responsible for racial disparities or that poverty derives largely from structural constraints. More women began to question the legitimacy of gender disparities but not the legitimacy of other social inequalities. The growing consciousness of gender inequity resulted primarily from shifts in viewing its cause, not the appropriateness of gender roles themselves. Women's evaluations of traditional and nontraditional gender roles remained quite stable. Slightly less than half in 1972 and only a bit more than half in 1976 and in 1983 said that men and women should have an equal role in running business, industry, and government. Women did not become more critical of gender differentiation itself but changed their views of its causes.

The Likert items, which separate blame of the system from blame of women, clarify the changes in their explanations of disparity. These separate measures of individual and structural attributions were asked in 1972 and 1983. In those 10 years women shifted especially from an explanation based on natural endowments. Far fewer women in 1983 believed that men are more qualified for highly responsible jobs or that men are born with more ambition and drive to succeed or that women are happiest making a home and caring for children. By 1983 different endowments and preferences no longer easily served as a basis for legitimacy. It is this shift that primarily accounts for less legitimacy for a system that allocates unequal job status and wages to men and women. The items that referred specifically to systemic obstacles—job discrimination, lack of educational opportunity—remained much the same in these years. Separating structural explanations from those that emphasize personal qualities, as the Likert items do, allows us to understand that women's views of legitimacy changed primarily because of their growing doubts about men's claims to privilege.

Differing responses to the questions about power exhibit the same pattern. Many women became critical of men's power and influence in life and politics. Of all the issues examined across these years, this issue shows the most dramatic change. In 1976, when the question about men's power was first asked, only a quarter of the women said men have "too much" power. This proportion nearly doubled by 1983. In the same period, women's evaluations of their own power showed only minor change. Women certainly felt more aggrieved about their power by 1983, but that resulted nearly entirely from their growing feeling that men as a category are too powerful in the *public* arena. (These questions do not probe the private world of informal and intimate relationships between women and men, nor do they pertain to individual men but to men as a category.)

These differences and stabilities tell a clear story. In the aggregate, women became more politically conscious in these years, but this primarily reflected changes in their opinions toward men's power and natural endowments for economic success. Women changed their appraisals of women's position in society a bit, but more strikingly, they reappraised the position of men. For many women, men as a category had lost some of their special claims to economic superiority, privilege, and power.

SUBGROUP SHIFTS

The panel of 742 women followed from 1972 to 1976 was used to assess which groups of women changed most and least. We carried out four regression analyses in which the statuses women held in 1972 were treated as predictors of the 1976 indexes of identification, discontent, withdrawal of legitimacy, and collective orientation. In these analyses, the 1976 scores on gender consciousness were also adjusted for the women's initial positions on the same dimensions. For example, 1976 discontent scores were predicted from 1972 discontent and the following 1972 statuses: amount of education, labor force status, marital status, and age. The question is whether certain statuses in 1972 set the stage for some women to become more politically conscious than others.

These regression analyses showed that women with more education and younger women became significantly more identified, discontented, and critical of legitimacy, and that marital and labor force status had less influence on change. Working in 1972 did result in becoming significantly more critical of the legitimacy of market disparities, and discontent grew significantly more among women who were unmarried in 1972.[11]

---

[11] The effects of *status changes* between 1972 and 1976 are analyzed in detail elsewhere. Labor force changes proved consistently more important than marital changes, acquiring more schooling, or shifts in religious attendance (Gurin, 1984).

Table 2 gives the 1972–1976 comparisons between various subgroups of women where the regression analyses revealed statistically significant effects of 1972 statuses. It shows that college-educated women were more politically conscious both in 1972 and 1976 and that their growth in consciousness was also greater than that of other women. In particular, they became much more critical of the legitimacy of wage and other market disparities, as did women who were already employed in 1972. Change was also greater among younger women, especially those under 30 when the study began. Thus, despite "less room to move," the greatest shifts occurred among subgroups who were originally the most politically conscious.

PARALLELISM OF SHIFTS FOR MEN AND WOMEN

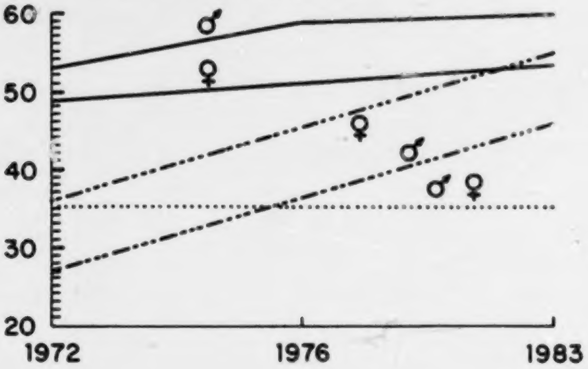
Men moved right along with women and their parallel shifts are striking (Figure 2). Neither group changed in attitudes toward action women ought to take. Both also maintained nearly the same opinions toward equal roles that they held in 1972. Both increasingly adopted a structural view of disparity, which was particularly apparent in the larger

Table 2. Endorsements of Gender Consciousness in 1972 and 1976[a] of Women in Different Statuses in 1972

								Withdrawal of Legitimacy		
		Identification			*Discontent*			*Scores of 4 or 5 on index where 5 = attribution to structural forces*		
		Chose women as category to which felt "closest"			*Said women have "too little power"*					
	(N)	*'72*	*'76*	*% Diff.*	*'72*	*'76*	*% Diff.*	*'72*	*'76*	*% Diff.*
Education										
Less than H.S.	(251)	7%	7%	(0)	31%	30%	(0)	30%	37%	(7)
High School	(263)	10%	10%	(0)	36%	42%	(6)	46%	56%	(10)
College	(227)	12%	19%	(7)	37%	47%	(10)	53%	77%	(24)
Age										
Under 30	(175)	12%	19%	(7)	44%	53%	(11)	63%	78%	(15)
30-49	(272)	13%	13%	(0)	39%	44%	(5)	48%	63%	(15)
50-69	(218)	4%	4%	(0)	29%	29%	(0)	29%	39%	(10)
70+	(71)	3%	5%	(2)	16%	21%	(5)	23%	31%	(7)
Labor force										
Not in labor force	(410)							39%	47%	(8)
In labor force	(332)							48%	69%	(21)
Marital status										
Married	(603)				33%	36%	(3)			
Not married	(139)				44%	55%	(11)			

[a] The only percentages shown in this table are for those 1972 statuses that revealed statistically significant effects in the regression analyses. There were no significant effects on Collective Orientation.

2a. Withdrawal of Legitimacy

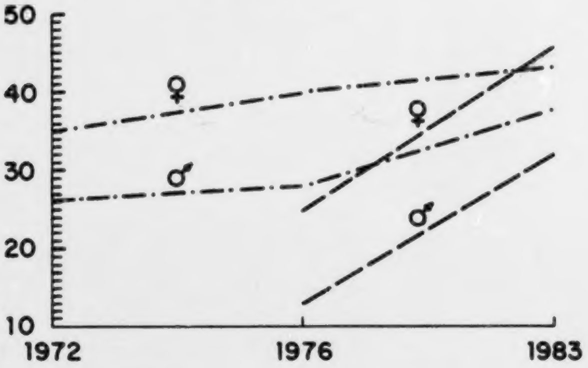


——— Advocate = Roles (endorsed 3 most extreme positions on 7-point scale). Males higher 1976, 1983, $p = .01$

—···— Reject natural endowments explanation disparity (scores 4-5 on 5-point index). Females higher 1972, 1983, $p = .001$

·········· Accept system explanation (scores 4-5 on 5-point index).

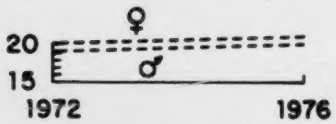
2b. Collective Discontent



·—·— Women have "too little power." Females higher 1972, 1976, $p = .01$

——— Men have "too much power." Females higher 1976, 1983, $p = .001$

2c. Collective Orientation



------ Advocate collective strategies (scores 4-5 on 5-point index).

Figure 2. Male ♂ and Female ♀ Trends

numbers in 1983 who discounted men's superior qualifications and women's tastes for homemaking as explanations for wage and labor status disparities. Men, like women, also changed their evaluations of the power each ought to have. Men shifted their appraisals of women's power so much that by 1983 the two groups no longer differed significantly. Even more impressive, men greatly changed evaluations of their own power. By 1983 more men dramatically conceded that men have "too much" power. Over and over it is clear that men and women were changing together in these years. Women became more politically conscious and men became more sympathetic to a feminist analysis of gender disparity.

One further point should be noted about these shifts. There has been no backlash or backward trend in either women's or men's opinions. Goode (1982) also stresses this in examining data collected between the 1940s and 1970s on men's opinions toward women's progress toward equality. Many have speculated, however, that the increased conservatism in the country might have produced some retrogression after 1980. We find none, despite clear evidence that both men and women were less liberal politically in 1983. Three indicators of liberalism and conservatism typically asked in the national election studies were repeated in 1983 to put the shifts in gender consciousness in broader political perspective. These are for self-placement on the liberal-conservative scale, attitude toward the government guaranteeing jobs and a good standard of living, and attitude toward government aid to minorities. On all three scales, fewer people endorsed the liberal position in 1983 than in 1972. Despite this, more rather than fewer people took a feminist stance toward gender relations in 1983.

Conclusions

Several conclusions can be drawn about women's gender consciousness as defined and measured here.

First, in most respects women's gender consciousness has been comparatively weak. Structural constraints that should operate against the development of women's strong political consciousness have apparently done so. The relative weakness is demonstrated in two ways. Except for their strong criticism of the legitimacy of gender disparities in the labor market, women's political consciousness was weaker than that of blacks, older people, or blue-collar workers. Moreover, women did not express a distinctively subordinate consciousness, and this contrasted dramatically with the subordinate-superordinate polarization of the perspectives of blacks and whites.

Second, although their consciousness was weak in this comparative sense, women did become more politically conscious over the 1970s. The shifts were impressive on two components of consciousness, discontent and withdrawal of legitimacy, and especially among women who were under 30, college-educated, employed, and unmarried when we began monitoring trends in 1972.

Third, men changed along with women. They increasingly discounted their superiority and claims of men to power and privilege, and recognized women's relative deprivation of power and influence in life and politics. This parallelism, as well as the continuing structural interdependence of women and men and the growing numbers of employed women, suggest that gender relations will go on changing, without reaching the level of racial conflict that has occurred in this country and the religious and ethnic cleavages elsewhere in the world. As women change their opinions toward gender relations, at least partly because of employment and experience in various nontraditional statuses (Gurin, 1984), men will change too. This is not to underestimate the importance of the women's movement and the need to mobilize women to exert pressure for change. Still, these data make clear that men are far more supportive of gender equality than whites are of racial equality, a conclusion also drawn by Smith and Kluegel (1984). Blacks cannot depend as much on support from superordinates in their struggle for equality because blacks and whites have very disparate views of the race situation. Instead, blacks' group consciousness, which is fostered by the structure of race relations, remains their major resource for mobilizing and effecting change. While the structure of gender relations, compared to that of race relations, inhibits the development of strong group consciousness among women, it simultaneously provides a different avenue for social change—one that capitalizes on the greater consensus among men and women.

Fourth, the gender patterns do suggest some tension in gender relations that should not be obscured by the relative consensus of men and women. Although both changed in similar ways between 1972 and 1983, two critical differences stand out in the trends shown in Figure 2. Men still differed from women in 1983 in appraising men's power and the legitimacy of their market superiority, just as they did in 1972. True, larger numbers began to question men's special claims to leadership, economic success, and political power. But in 1983 they still discounted these claims less than women did. The justice of men's advantage is at the center of both change and continuing disagreement between the sexes. To the extent that a gender cleavage exists, it is about men's privilege. In the eyes of both men and women, the legitimacy of men's advantage has slipped, but even now more men maintain that they deserve their relative economic and political advantage.

References

Andersen, K.
1975 "Working women and political participation, 1952–1972." American Journal of Political Science 19:439–53.

Baxter, S., and M. Lansing
1983 Women and Politics: The Visible Majority (rev. ed.). Ann Arbor: The University of Michigan Press.

Bentler, P. M., and D. G. Bonett
1980 "Significance tests and goodness of fit in the analysis of covariance structures." Psychological Bulletin 88:588–606.

Blydenburgh, J. C., and R. S. Sigel
1983 "Key factors in the 1982 elections as seen by men and women voters: an exploration into the vulnerability thesis." Paper delivered at the 1983 Annual Meeting of the American Political Science Association, Chicago.

Brickman, P., R. Folger, E. Goode, and Y. Schul
1980 "Micro and macro justice." In M. J. Lerner (ed.), The Justice Motive in Social Behavior. New York: Plenum.

Commission of the European Communities
1984 European Men and Women in 1983. Brussels.

Coser, L. A.
1956 The Functions of Social Conflict. New York: Free Press.

Crosby, F. J.
1982 Relative Deprivation and Working Women. New York: Oxford University Press.

Dawes, R. M., and B. Corrigan
1974 "Linear models and decision-making." Psychological Bulletin, 81:95–106.

Epstein, C. F.
1984 "Ideal images and real roles." Dissent (Fall):441–52.

Fireman, B., and W. A. Gamson.
1979 "Utilitarian logic in the resource mobilization perspective." In M. Zald and J. McCarthy (eds.), The Dynamics of Social Movements. Cambridge, MA: Winthrop.

Frankovic, K. A.
1982 "Sex and politics—new alignments old issues." PS, 15:439–48.

Goode, W. J.
1982 "Why men resist." In B. Thornet and M. Yalom (eds.), Rethinking the Family: Some Feminist Questions. New York: Longmans.

Gurin, P.
1984 "Gender consciousness: the impact of status change." Unpublished paper, Russell Sage Foundation, New York.

Gurin, P., H. Miller, and G. Gurin
1980 "Stratum identification and consciousness." Social Psychology Quarterly 43: 30–47.

Jennings, M. K., and B. G. Farah
1980 "Gender and politics: convergence or differentiation?" Paper presented at a June conference on political action, Bellagio, Italy.

Lake, C. C.
1982 "Guns, butter, and equality: the women's vote in 1980." Paper presented at the Midwest Political Science Association meetings, April 28–May 1, Milwaukee, WI.

Lane, R. E.
1962 Political Ideology: Why the American Common Man Believes What He Does. New York: The Free Press.

Lerner, M. J.
1981 The Belief in a Just World. New York: Plenum.

Mannheim, K.
1952 The Problem of Generations: Essays on the Sociology of Knowledge. London: Routledge and Kegan Paul.
Merton, R. K.
1968 Social Theory and Social Structure. New York: The Free Press.
Merton, R. K., and A. S. Kitt [Rossi]
1950 "Contributions to the theory of reference group behavior." In R. K. Merton and P. F. Lazarsfeld (eds.), Continuities in Social Research: Studies in the Scope and Method of the American Soldier. Glencoe, IL: The Free Press.
Miller, A. H., P. Gurin, G. Gurin, and O. Malanchuk
1981 "Group consciousness and political participation." American Journal of Political Science 25:494–511.
Oberschall, A.
1978 "Theories of social conflict." Annual Review of Sociology 4:291–315.
Pomper, G. M.
1975 Voters' Choice: Varieties of American Electoral Behavior. New York: Dodd, Mead.
Scientific Manpower Commission
1983 Professional Women and Minorities: A Manpower Data Resource Service. Washington, D. C.
Shapiro, R., H. Mahajan, and K. Veith
1984 "Gender differences in policy choices: trends from the 1960s to the 1980s." Paper prepared for presentation at meeting of the Midwest Association for Public Opinion Research, November 16–17, Chicago, IL.
Smith, E. R., and J. R. Kleugel
1984 "Beliefs and attitudes about women's opportunity: comparisons with beliefs about blacks and a general perspective." Social Psychology Quarterly 47:81–95.
Smith, T. W.
1984 "The polls: gender and attitudes toward violence." Public Opinion Quarterly 48:384–96.
Stephan, W. G.
In Press "Intergroup relations." In G. Lindzey and E. Aronson (eds.), The Handbook of Social Psychology, forthcoming.
Tajfel, H.
1978 "Social categorization, social identity and social comparison." In H. Tajfel (ed.), Differentiation Between Social Groups. New York: Academic Press.
Thornton, A., D. F. Alwin, and D. Camburn
1983 "Causes and consequences of sex-role attitudes and attitude change." American Sociological Review 48:211–27.
Thornton, A., and D. Freedman
1979 "Changes in the sex-role attitudes of women, 1962–1977: evidence from a panel study." American Sociological Review 44:831–42.
Tilly, C.
1978 From Mobilization to Revolution. Reading, MA: Addison Wesley.
Verba, S., and N. H. Nie
1972 Participation in America. New York: Harper and Row.
Welch, S.
1977 "Women as political animals? A test of some explanations for male-female political participation differences." American Journal of Political Science 21:711–30.
Williams, R. M., Jr.
1975 "Relative deprivation." In L. A. Coser (ed.), The Idea of Social Structure: Papers in Honor of Robert K. Merton. New York: Harcourt Brace.

Myth and Reality: The ERA and the Gender Gap in the 1980 Election

JANE J. MANSBRIDGE

ALMOST all feminist political activists believe that women's support for the Equal Rights Amendment (ERA) played a major role in producing the gender gap in the 1980 presidential election. This view was prompted by a virtual consensus among news analysts after the election, and bolstered by problematic data from the political science profession.

This paper argues that the gender differences in 1980 presidential voting were not due to attitudes toward the ERA. It then asks why the "ERA explanation" of the gender gap persisted, despite much evidence against it.

The analysis relies primarily on data from the CBS News/New York Times 1980 Election Day Survey (hereafter called the 1980 exit poll).

Abstract Attitudes toward the Equal Rights Amendment were not an important factor in the emergence of the "gender gap" in the 1980 presidential election. Conclusions to the contrary by news analysts, feminist political activists, and political scientists are based on a combination of the power of expectation, faulty analysis, and random bias in the most frequently used survey in political science.

Jane J. Mansbridge is Associate Professor of Political Science and Sociology, and Research Faculty at the Center for Urban Affairs and Policy Research, Northwestern University. The author wishes to thank Gary Winters for computer work, and Ronald King, Steven Jackson, Fay Lomax Cook, and Christopher Jencks for comments on an earlier version of this paper, presented at the 1983 Annual Meeting of the Midwest Political Science Association in Chicago. Kathleen Frankovic of CBS News was extremely helpful in providing CBS/New York Times data before they became widely available.

The data analyzed herein were collected by the New York Times and CBS News and processed by the New York Times and CBS News under a grant from the Russell Sage Foundation; by Warren E. Miller and the National Election Studies of the Center for Political Studies at the University of Michigan; and by the National Opinion Research Center of the University of Chicago. The data were provided through the Interuniversity Consortium for Political and Social Research of the University of Michigan. The New York Times, CBS News, the CPS, NORC, and the ICPSR bear no responsibility for the analysis and interpretations presented here.

Public Opinion Quarterly Vol. 49:164–178 © by the Trustees of Columbia University
Published by Elsevier Science Publishing Co., Inc. 0033-362X/85/0049-164/$2.50

This survey almost certainly provides the best available data on the 1980 election. It was conducted among voters as they left the polls, not at some later date, and the sample (N = 15,201) was far larger than any other study.[1] When voters were handed a questionnaire, the front side included the following question:

In the presidential election who did you just vote for?
1. Jimmy Carter
2. Ronald Reagan
3. John Anderson
4. Ed Clark
5. Barry Commoner
6. Didn't vote for President
7. Other

On the back, the following question appeared:

Please mark an "X" to show if you agree or disagree with each of the following statements: . . . I support the Equal Rights Amendment—ERA—the constitutional amendment concerning women.

Of the unweighted sample of 15,201, 2452 did not answer the ERA question, but did answer one or more other questions on the back of the questionnaire. These responses have been coded "don't know." Another 975 respondents did not answer *any* questions on the back of the questionnaire. These respondents seemingly did not turn the questionnaire over and have been eliminated from the analysis.

Gender Differences

In the 1980 presidential election, only 47 percent of female voters supported Ronald Reagan, compared to 56 percent of male voters—a difference, or "gender gap," of 9 percent.[2] This gender difference may not have been much greater than the gender difference in 1972 (Poole and

---

[1] The precincts were a random subsample of those used by CBS News for estimating the election outcome. The sampling frame consisted of all precincts within a state, stratified by party vote and geography. Sample precincts were selected with probability proportionate to the total vote cast in a recent election. Within precincts, respondents were selected on a systematic random basis, with the interviewer having no control over respondent selection. The unweighted sample of 15,201 cases was subsequently weighted, taking into account the probability of selection, a noninterviewer adjustment, and a ratio estimate to the final vote totals. All weights were reduced by a constant so that the weighted sample approximated the unweighted sample in size.

[2] CBS/New York Times exit poll. For comparative purposes, the 1980 CPS American National Election Study (NES) produces a retrospective gender gap of 7 percentage points (48% of women voting for Reagan compared to 55% of men) while the 1982 NORC General Social Survey (GSS) produces a retrospective gender gap of only 4 percentage points (45% of women voting for Reagan compared to 49% of men).

Zeigler, 1982). Nonetheless, Reagan's announced opposition to the Equal Rights Amendment during the campaign and the publicity given to this and Reagan's other "anti-woman" positions by the newly powerful National Organization for Women (NOW) made the gender gap salient in a way it had never been before. Indeed, it seems to have been NOW that coined the term "gender gap," and it was certainly NOW that brought the term into general currency.

Were These Differences Due to the ERA?

The theory that differences in men's and women's attitudes toward the ERA contributed substantially to the gender gap is certainly plausible on its face. The main problem with the theory is that there was very little difference in men's and women's attitudes toward the ERA in 1980. Table I shows that women voters were only 3.7 percent more likely to favor the ERA than men.[3] Thus, even if *all* pro-ERA voters had voted against Reagan and vice versa, the ERA could only have accounted for about a third of the gender gap in 1980. Since, as we shall see, most pro-ERA voters opposed Reagan for a multitude of reasons besides his ERA stand, Table I suggests that the ERA's contribution to the gender gap was miniscule.

Table I could be misleading, however, if women felt more intensely about the ERA than men, and if women's feelings about the ERA therefore affected their vote more than men's feelings did. Although the evidence is mixed on the degree to which men and women differed in the intensity of their support for the ERA, the crucial issue for present

Table 1. ERA Support by Gender, 1980

	Women	*Men*
Favor	47.6	43.9
Oppose	34.2	39.8
Don't know	18.2	16.3
(N)	(7035)	(7182)

SOURCE: New York Times/CBS News exit poll, Nov. 4, 1980.

---

[3] Although men's support for the ERA seems to have diminished to a slight degree over time while women's has slightly increased, the differences are very small, and polls from the same time period can report them in opposite directions. For example, although the GSS for 1982 reports 70 percent of the women surveyed supporting the ERA, compared to 67 percent of the men, a CBS News/New York Times poll from the same year reports (for a question producing lower overall support) 55 percent of the men surveyed supporting the Equal Rights Amendment, compared to 53 percent of the women (Clymer, 1982, no survey date, question wording, or *N* given).

Table 2. Vote in 1980 Election by ERA Stance and Gender

	Women			*Men*		
	pro-ERA	*anti-ERA*	*Difference*	*pro-ERA*	*anti-ERA*	*Difference*
Reagan	35%	65%	−30	41%	73%	−32
Carter	53	30	+23	46	23	+23
Anderson	11	4	+7	11	4	+7
Other	2	1	+1	2	1	+1
(N)	(3342)	(2401)		(3145)	(2843)	

SOURCE: New York Times/CBS News exit poll, Nov. 4, 1980.

purposes is not intensity per se, but whether support for the ERA affected women's *voting behavior* more than it affected men's.

When we look at the way men and women actually voted in 1980, we find no evidence that women's views on the ERA influenced their voting behavior more than men's views on the ERA did. While widespread citizen support for the ERA may have hurt Reagan,[4] Table 2 indicates that it hurt him equally with both men and women. The correlation between supporting the ERA and voting against Reagan (the only anti-ERA candidate) was 0.304 among women ($N = 5743$) and 0.315 among men ($N = 5989$).[5] Thus, although women were more likely than men to indicate on some surveys (including this one) that the ERA was an important issue, women were no more likely than men to *vote* for a candidate who agreed with them about the ERA. Since men and women were almost equally likely to favor the ERA, and were equally likely to vote for candidates who shared their views on the ERA, there is no way the ERA could explain much of the "gender gap" in 1980.

---

[4] The observed correlation between attitudes toward the ERA and presidential preference in 1980 was higher than the correlation between most other attitudinal measures and presidential preference. Indeed, Poole and Zeigler's (forthcoming) analysis of the 1980 NES data indicates (Table 2.10) that one's opinion on the ERA predicted one's vote in the Reagan/Carter election better than one's opinion on any other measured issue except government spending (the best predictor) and whether or not the government should guarantee jobs and a good standard of living.

The correlation between supporting the ERA and opposing Reagan was 0.313 in the CBS/NY Times exit poll, and 0.305 in the 1980 NES. We know, however, that support for the ERA was correlated both with "liberal" attitudes on many other issues and with party identification. When we add measures of liberalism/conservatism, party identification, and a set of demographic variables to an equation predicting whether people voted for Reagan, the standardized coefficient falls to 0.166 in the New York Times/CBS poll, and to 0.134 in the NES. When we add other attitudinal variables, available only in the NES, it falls to 0.089. This tells us that at least 70 percent of the observed correlation between favoring the ERA and opposing Reagan is spurious. (Party identification and liberalism/conservatism together account for most of this change. With liberalism/conservatism and party identification controlled, the standardized coefficient of ERA in the NES is 0.122 for women compared to 0.125 for men. Since the dependent variable is dichotomous, purists can rightly argue that least-squares regression is inappropriate. But since the split is about 50-50, OLS results are not subject to serious bias, and they have the virtue of being easy to compute and interpret.)

If even 30 percent of the observed relationship were truly causal, the ERA would still have had a substantial effect on the outcome of the election. If we look at unstandardized coefficients, we find that even with liberalism, party identification, and other demographic and attitudinal variables in the NES controlled, the percentage of strongly pro-ERA voters opposing Reagan was about 15 points higher than the percentage of anti-ERA voters opposing him.

The difficulty with this interpretation is, of course, that the kinds of demographic and attitudinal measures available on most surveys do not capture all the potentially relevant differences between the ERA supporters and opponents. The most that can be said, therefore, is that while Reagan's opposition to the ERA probably cost him some votes in 1980, it is not clear how many.

What, then, was the cause of the gender gap in 1980? As Kathleen Frankovic (1982) has pointed out, that gender gap was primarily explained by men's and women's differing attitudes toward the risk of war. One of the "agree-disagree" questions on the CBS/New York Times 1980 exit poll was:

> We should be more forceful in our dealings with the Soviet Union even if it increases the risk of war.

Frankovic demonstrates that women were much more likely than men to disagree with this statement, and that with attitudes on this question controlled, the gender gap in the 1980 exit poll fell to 3 percent. The reduction was "effective in all age groups and at all educational levels."

A full regression analysis of the CBS/New York Times data allows us to quantify the importance of most of the factors that contributed to the gender gap in 1980. Table 3 lists all the variables that account for more than 2.5 percent of the gender gap.[6] Column 1 shows the correlation of each variable with the gender variable, Female. Column 2 shows the standardized regression coefficient of each variable in an equation predicting whether the respondent voted against Reagan. Column 3 shows the size of the gender gap that would have arisen if the variable in question had been the only factor contributing to the gender gap. (Column 3 sums

---

[5] Since these correlation coefficients involve dichotomous variables, they are subject to the usual caveats. They do, however, provide a reasonable basis for comparing the strength of the association among males to that among females, since the marginals for males and females differ by roughly comparable amounts. Readers who prefer some other measure of association can calculate it from Table 2. The correlations, like the figures in Table 2, use the weighted sample and are calculated only for voters expressing an opinion on the ERA and voting for president. Coding the ERA variable so that "don't know" responses fall between "agree" and "disagree" does not affect the results.

[6] All variables are dummy variables except Favors ERA, which is coded so that the "don't know" responses fall between "agree" and "disagree," in order to maximize cases. Coding to eliminate these responses does not affect the results. For party identification the omitted category is Republican and no answer; for U.S. House vote the omitted category is "some other party" and no answer. In the regression, anti-Reagan vote is coded 1, Reagan vote 0.

Table 3. Variables Contributing to Gender Differences in Opposition to Reagan: 1980

Variable	*Correlation with Female*	*Standardized Coefficient in Anti-Reagan Equation*	*Contribution to Gender Gap:* *Absolute*[a]	*%*[b]	*Grouped %*
Favors ERA	.0509	.0805	.0041	4.7	9.5
Favors ERA—ERA/ Abortion most important issue	.0709	.0627	.0045	5.1	
Favors ERA—ERA/ Abortion next most important	.0722	.0323	.0023	2.7	
Opposes ERA—ERA/ Abortion most important issue	.0324	−.0374	−.0012	−1.4	
Opposes ERA—ERA/ Abortion next most important	.0407	−.0336	−.0014	−1.6	
Oppose more forceful dealings with USSR if risk of war	.1685	.1352	.0228	26.3	26.3
Usually think of self as Democrat	.0611	.2687	.0164	18.9	25.6
Usually think of self as Independent	−.0408	.1334	−.0054	−6.3	
Voted for Democratic candidate for U.S. House	.0254	.1623	.0041	4.7	
Voted for Republican candidate for U.S. House	−.0610	−.1185	.0072	8.3	
Interviewer observation of race—Black	.0579	.1052	.0061	7.0	25.6
Crisis in Iran most or next most important issue	.0528	.0615	.0035	3.7	
Balancing the federal budget most or next most important issue	−.0531	−.0594	.0032	3.6	
U.S. prestige around the world most or next most important issue	−.0546	−.0428	.0023	2.6	
Inflation and economy most or next most important issue	−.0301	−.0735	.0022	2.6	
Professional occupational status	−.0597	−.0367	.0022	2.5	
Miscellaneous[c]	—	—	.0048	3.6	
Unexplained	—	—	.0113	13.0	13.0
Total			.0867	100.0	100.0

SOURCE: New York Times/CBS News exit poll, Nov. 4, 1980.

[a] Calculated as $B_{XF}B_{RX}$, where B_{XF} is the unstandardized bivariate regression of the variable on Female, and B_{RX} is the unstandardized regression of Anti-Reagan vote on the variable, with all the other variables in the table controlled.

[b] $B_{XF}B_{RX}/B_{RF}$, where B_{RF} is the unstandardized regression of Anti-Reagan vote on Female.

[c] Includes, in order of importance, thinking tax cuts more important than a balanced budget, liberal-conservative identification, thinking unemployment a greater problem than inflation, importance of cities and taxes as issues, region, rural residence, age, religion, pro-Kennedy sentiment, Hispanic identification, and financial situation.

to .0867, which is the total gender gap.) Column 4 shows the percentage of the actual gender gap (.0867) accounted for by this variable.

Among the 14,185 respondents with complete data, 8.7 percent more women than men voted for Reagan. Variables that include or mention the ERA account for 9.5 percent of this gender gap. Thus if the ERA-related variables had been the only source of the gender gap, only 0.8 percent more women than men would have voted against Reagan. Other demographic and attitudinal variables measured in this survey—including the question on the Soviet Union, which had the greatest effect, and usually thinking of oneself as a Democrat, which had the next greatest effect—account for another 77.5 percent of the gap, leaving 13 percent unexplained.[7]

The ERA-related variables fall into two categories: a simple measure of whether the respondent favored or opposed the ERA, and a set of variables, equivalent to interaction terms, that combine the respondent's attitude toward the ERA with measures of whether the respondent mentioned "ERA/ Abortion" either first or second in response to the question, "Which issues were most important in deciding how you voted today?" The dichotomous measure of support or opposition to the ERA produces a gender gap of 0.4 percent. The interaction between attitudes toward the ERA and intensity regarding either the ERA or abortion accounts for another 0.4 percent. Including intensity regarding abortion as part of the "ERA-related variables" almost certainly leads to an overestimate of the effect of intensity.

Why the Misperception?

As Lindblom and Cohen (1980) have pointed out, we are most likely to pick up and use social science data that confirm our expectations. The natural expectation that a heavily publicized women's issue like the ERA would be likely to create a gender gap produced a series of faulty readings of the survey data. These faulty readings began with the "crash analyses" of reporters under a deadline, spread to a feminist press inexperienced in survey research, and received professional confirmation through a typographical error in one professional source and reliance on statistically insignificant results in another. From this point on, feminist activists could support their prior beliefs with professionally sanctified data.

---

[7] One might conjecture that attitudes toward the ERA either affected one's party identification and liberal-conservative identification, and/or affected one's vote in the 1980 election, which consequently affected one's party identification and liberal-conservative identification. If this were true, and party and liberal-conservative identification had no independent effect on the vote (i.e., were omitted from the regression equation), the combined effect of all ERA-related variables would explain 13.3 percent, as opposed to 9.5 percent of the total gender gap.

The story begins four days after the 1980 presidential election, when, on the basis of a "crash" background analysis of the CBS/New York Times News exit poll by Edward Tufte,[8] Adam Clymer of *The New York Times* suggested that "his opposition to the Equal Rights Amendment handicapped Mr. Reagan's bid for their [women voters'] support." In support of this contention, Clymer presented an earlier version of the same data that appear in the first two columns of Table 2, showing that women who opposed the Equal Rights Amendment were far more likely to vote for Reagan than women who favored the Equal Rights Amendment (Clymer, 1980). Unfortunately, Clymer did not look at analogous data for men, which showed exactly the same pattern.

The feminist press immediately picked up the *New York Times* analysis. Indeed, the National Organization for Women put out a 20-page booklet, entitled *Women Can Make the Difference*, devoted entirely to the gender gap in the polls. This pamphlet (NOW, 1981) recast Clymer's figures into a table, repeated his analysis, and concluded that although "many analysts believed" that the gender gap in 1980 derived from differences between men and women in the fear of war, the New York Times/CBS exit poll indicated that "female voting behavior was also based on Reagan's anti-women's rights positions, especially the Equal Rights Amendment."[9] A similarly worded statement appeared in the *National NOW Times* (1980–81).

Clymer was certainly not alone in misreading the first cross-tabulations on gender, voting, and ERA support. G. Evans Witt (1981), director of polling for the Washington Bureau of Associated Press, wrote,

> Why do women feel this way about Reagan? One explanation, put forward last fall, argued that women are traditionally more pacifist than men. . . . But the election day survey by AP/NBC News found a stronger relationship between opinions on the ERA and voting in the presidential race than between attitudes on the "war and peace" issue and the vote for and against Reagan. The warmonger image was certainly a factor in the election, but for women the issue of women's rights was more significant.

Witt's factual premise was correct. There was indeed a stronger relationship between opinions on the ERA and voting choice than between attitudes on war and peace and voting choice. This reflects the fact that attitudes toward the ERA were more party-related than attitudes toward war and peace. But this general relationship says nothing about the gender gap. It says that *people* who are pro-ERA are more likely to vote against Reagan than people who are pro-peace, but not that *women* who are pro-ERA are more likely to vote against Reagan than *men* who are pro-ERA.

---

[8] Adam Clymer, conversation with the author, April 1983.

[9] In reproducing these data, the NOW pamphlet made one unimportant typographical error, reporting the gap between pro-ERA and anti-ERA Anderson women as +4, rather than +7, as indicated in the *New York Times*.

The Clymer-NOW figures differ slightly from figures in Table 2 because the former were based on a "crash analysis" of 12,782 cases, completed four days after the election. Table 2, like all other calculations using the New York Times/CBS poll in this paper, is based on the full 15,201 cases available on ICPSR data tape #7812, weighted. Similar but not identical figures in an earlier paper (Mansbridge, 1983) were supplied by Kathleeen Frankovic of CBS.

Both these two misinterpretations derive from failing to show that women's attitudes toward the ERA differed from men's. A third misinterpretation springs from a *post hoc, ergo propter hoc* analysis. An unsigned article in *Opinion Outlook* (1981) reported that "when the GOP dropped support for the ERA from its platform, support for Reagan among women voters dropped also, especially among women who work outside the home." This statement implies strongly that the GOP's dropping ERA caused the drop in women's support for Reagan. But other things that the convention highlighted about Reagan, including his more hawkish foreign policies, could as easily have led to the drop in women's support. The CBS/New York Times voting data indicate that this must have been the case.

The assumption that women cared more about the ERA than men, and thus that the ERA must have contributed to the gender gap was so strong that a newspaper reporter could even fail to "hear" when a survey analyst denied the connection. Take the interview that Tom Smith of the National Opinion Research Center gave to Richard Cattani of the *Christian Science Monitor* in March 1981. According to Smith, he said, first, that the gender gap in voting was caused primarily by men's and women's different attitudes toward war and violence, and second, that men and women did *not* differ particularly on economic issues or on "social women's issues" like the ERA and abortion.[10] Yet Cattani reported the interview as follows:

> "In the last election, we had the largest sex difference in history—a 10 percent edge for Carter among women," says Tom Smith, director of the National Opinion Research Center.
>
> The principal election factor among men and women was economic performance under Carter, Mr. Smith says. After that, war and peace was the biggest concern among women, followed by the ERA and possibly abortion.

Despite intuitive convictions to the contrary, the gender gap was largely traceable to gender-related differences in attitudes toward violence and war, while "the ERA and possibly abortion" had essentially no role in the matter, as Smith had actually said.

In the political science profession, similar assumptions about the obvious connection of the ERA and the gender gap combined with a random bias in the data collected by Michigan's 1980 American National Election Study (NES) to convince several scholars that women's stronger support for the ERA in 1980 helped account for their lower vote for Reagan.

---

[10] Tom Smith, conversation with the author, June 1983.

As we have seen, the massive 1980 CBS/New York Times exit poll showed virtually identical correlations between supporting the ERA and voting against Reagan among men and women. But smaller surveys showed sizable random differences between the two correlations. In the 1982 NORC General Social Survey (GSS), the correlations were .168 for women (N = 473) and .280 for men (N = 359). This difference is not significant and is in the "unexpected" direction, so no one has made much of it. But in the 1980 NES the correlation between voting against Reagan and support for the ERA was .337 among women (N = 412) and .310 among men (N = 341).[11] While this difference is neither statistically nor substantively significant, it is in the "expected" direction. The underlying random bias in the NES data thus encouraged the conclusion that on this issue there might be important differences between men and women. Moreover, Keith T. Poole and L. Harmon Ziegler (1982), while not themselves arguing that their numbers explained the gender gap, presented data on the ERA that, due to a typographical error, greatly exaggerated the small random difference between men and women in the 1980 NES.[12]

These scholarly problems made it easy for feminists to conclude that the ERA contributed to the 1980 gender gap. Ethel Klein (1984), for example, makes the ERA central to her analysis of the "women's vote" in the 1980 election. She combines the 1980 NES question on the ERA with questions from the NES on the respondent's attitudes towards women's equal role in general and the respondent's perception of the candidates' stands on women's equal role,[13] and shows that in the NES this index had more influence on women's voting behavior than on men's. She then concludes that:

---

[11] These NES correlation coefficients (which, like the GSS ones, involve dichotomous variables—see note 4) are calculated with only the validated plus the unchecked voters (NES variable 1207). Using either the full sample or the smaller sample of only validated voters increases the difference between the correlations, but not enough for the difference to reach statistical significance. Both the NES and GSS calculations exclude those who have no opinion on the ERA, and the GSS calculations exclude those who have never heard or read of the ERA. Including the no opinion responses by coding them to fall between support and opposition to the ERA does not affect the results.

[12] In Poole and Zeigler (1982, Table 10), the figure of 24.4 for men, which represented "the difference between Reagan and Carter vote on the issue (maximum possible value is 100)," has been corrected in Poole and Zeigler (forthcoming, Table 2.10) to read 30.4, as compared to a score of 33.2 for women.

[13] The exact wording of the equal role scale was: "Recently there has been a lot of talk about women's rights. Some people feel that women should have an equal role with men in running business, industry and government. Others feel that women's place is in the home.

> By 1980, women's rights was a campaign issue. Both men and women favored the ERA and saw Carter as the better candidate on feminist issues, but women voted for Carter because he lent women his support, while men did not. Although men were sympathetic to ERA, other issues were more important to them than women's rights were. Because women confront the problems wrought by sex inequality in their daily lives, women have consistently voted on the basis of promoting their rights ever since these issues were raised in the early 1970's (p. 164).

Unfortunately, Klein's analysis never tries to distinguish the effects of Reagan's opposition to the ERA from the effects of his perceived opposition to changes in women's traditional role as homemakers. Her data suggest, however, that the ERA per se had minimal effects, and that it was women's belief that Reagan opposed changes in their traditional role that cost him their support.[14]

On the basis of survey advice from Klein,[15] Eleanor Smeal, a former president of NOW, also concludes that the ERA was an important cause of the gender gap in the 1980 election in her recent book, *Why and How Women Will Elect the Next President* (Smeal, 1984; also Perlez, 1984). Although Smeal recognizes "the conventional wisdom" that "women and men were pro-ERA by roughly the same high support levels," she counters that "those of us who had spent the last decade working for women's rights issues knew better. Public opinion polls often do not measure intensity of opinion or correlate voting behavior with issue support" (p. 10).

Klein and Smeal are correct when they argue that women felt more strongly about the "women's issues" than men. As Smeal says, "When CBS/New York Times asked voters what were the two top issues for them, 11 percent of the surveyed women and 5 percent of the men ranked

---

Where would you place yourself on this scale or haven't you thought much about this? . . . Where would you place Jimmy Carter [Ronald Reagan, etc.] (on this scale)?"

[14] See Klein (1984), Figure 9.3 and Table 9.9. The male-female difference in the effects of the ERA variable is never significant in my analyses of the NES. Klein does not test the significance of the differences between males and females even for her combined variables, although she reports that some of her coefficients were significantly different from zero for women but not for men.

[15] According to Smeal, "Ethel Klein, an assistant professor of government at Harvard University, provided assistance in interpreting and analyzing polling data and in assembling research for footnotes in the first four chapters [those discussed here]. She reviewed all data in those chapters, checked their statistical accuracy and significance, and wrote all the footnotes concerning polling data for these chapters" ((Smeal, 1984:184).

ERA/abortion as being most important" (p. 20). What Smeal neglects to say is that for many women it was abortion, not the ERA, that really mattered. She also neglects to say that a large minority of the women who thought the ERA and abortion important were *opposed* to one or both. Thus, as Table 3 shows, the *net* effect of the fact that women felt more strongly about the ERA and abortion was to widen the gender gap by only 0.4 precent.

It was also true, as Smeal claimed, that 90 percent of the ERA volunteers and activists were women" (p. 10). But activists are a tiny minority of the electorate, and survey data show that activists' views on the ERA, like their views on many other subjects, were far from typical. If we combine the 1976, 1978, and 1980 NES surveys and the 1980 GSS, we find that 70 percent of the politically active women supported the ERA, compared to only 58 percent of the politically active men. But among the less active majority, the relationship disappears, with 72 percent of the men supporting the ERA, compared to 71 percent of the women.[16] The same pattern recurs when we look at those quintessential political activists, state legislators. In this group, three-quarters of the women supported the ERA, compared to only half the men (Stanwyck and Kleeman, 1983).[17]

---

[16] Because there are so few activists in small surveys, the differences between men and women activists in any one survey do not reach statistical significance, so one must pool several surveys in order to get reliable estimates. In the 1976, 1978, and 1980 NES, politically active men and women were defined as those who reported having engaged in four or more of the following: (1) voted in the most recent election, (2) tried to influence another's vote, (3) attended a political meeting, (4) worked on a campaign, (5) wore a campaign button or put a bumper sticker on one's car, and (6) gave money to a political campaign (in 1980, to a candidate, party, or PAC). Of the 146 active men in these surveys, 60 percent favored the ERA, compared to 69 percent of the 143 active women. Of the 1876 inactive or average men, 71 percent favored the ERA, compared to 70 percent of the 2381 inactive or average men. In the 1982 GSS, activists were defined as those who reported (1) voting in the 1976 presidential election, (2) voting in the 1980 presidential election, and (3) following what is going on in government and public affairs "most of the time." Of the 186 active men on this survey, 57 percent favored the ERA, compared to 71 percent of the 182 active women. Of the 342 inactive or average men, 78 percent favored the ERA, compared to 76 percent of the 523 inactive or average women. When we combine all four samples, the difference between active men and women is significant. In the 1980 exit poll, the difference is substantively small and not statistically significant, perhaps because the only measure of political activism is whether or not the respondent checked a box indicating that he or she was "active in politics."

[17] The exact figures are 77 percent of the female legislators ($N = 499$) agreeing that "The Equal Rights Amendment should be ratified," compared to 49 percent of the male legislators ($N = 267$). These figures derive from a mail-back questionnaire to all female state legislators and a systematic sample of male legislators, with a 57 percent response rate among the women, and a 52 percent response rate among the men. The "strongly agree" responses show an even greater gender difference: 62 percent of the women legislators, but only 26 percent of the men. I am grateful to Wendy Strimling, of the Center for the

Because the idea that the ERA had a major effect on the gender gap squared with what most of us expect, and because it corresponded with the actual experience of the politically active, the myth of the ERA's effect on the 1980 gender gap has persisted, both in scholarly works (cf. Traugott, 1982:89; Ladd, 1982:96) and in the feminist press (Riddlough, 1981; Walsh, 1983; Van Gelder, 1984; Eisenstein, 1984).[18]

Conclusion

The history of misinterpretation surrounding these data reads like a cautionary tale in an introductory textbook. We can draw the general morals that statistically insignificant differences are often due to chance, that reading only one half of a table is often misleading, and that media survey specialists making quick analyses against a deadline are likely to report results that confirm what they already believe is true. For actual politics, however, the lesson is more somber. An accurate understanding of the potential for feminist politics in this country would look carefully at the actual bases for popular support of specific women's issues. Listening primarily to activists will tend to produce an inaccurate picture of the population as a whole and will, inevitably, lead to mistaken political decisions.

References

Cattani, Richard J.
1981 "Reagan's standing with women still shaky." The Christian Science Monitor, July 9, p. 1.

Clymer, Adam
1980 "Displeasure with Carter turned many to Reagan." The New York Times, Nov. 9, I, p. 28.
1982 "Women's political habits show sharp change." The New York Times, June 30, p. 1.

American Women and Politics, for providing the data from which I could make these calculations. According to Stanwick and Kleeman, in the 15 states which failed to ratify the ERA by 1982, the gap was wider yet, with 76 percent of the female legislators supporting ratification, compared to 36 percent of the male legislators (no *N*'s given).

[18] This paper deals only with the gender gap in the 1980 presidential election. Lake's (1982) analysis of this election is consistent with the present analysis. For the gender gap in attitudes and in other elections, see Klein (1984)), Smeal (1984), as well as Jennings and Farah (1980), Lynn (1984), Miller and Malanchuk (1983), and Deitch (1984).

Deitch, Cynthia
1984 "The gender gap in social welfare ideology." Paper presented at the meetings of the Eastern Sociological Association, Boston MA, revised version forthcoming in Carol Mueller, The Politics of the Gender Gap. Beverly Hills, CA: Sage, 1986.

Eisenstein, Zillah
1984 "Contradictions inevitable as women become political force." In These Times, June 13–26, p. 25.

Frankovic, Kathleen A.
1982 "Sex and politics––new alignments, old issues." PS 15 (Summer) 439–48.

Jennings, M. Kent, and Barbara Farah
1980 "Gender and politics: convergence or differentiation." Presented at the Political Action Conference, Bellagio, Italy.

Klein, Ethel
1984 Consciousness and Politics: The Rise of Contemporary Feminism. Cambridge, MA: Harvard University Press.

Ladd, Everett Carll
1982 Where Have All the Voters Gone? 2nd ed. New York: W. W. Norton.

Lake, Celinda C.
1982 "Guns, butter and equality: the women's vote in 1980." Paper presented at the meetings of the Midwest Political Science Association, Milwaukee, WI.

Lindblom, Charles E., and David K. Cohen
1979 Usable Knowledge. New Haven: Yale University Press.

Lynn, Naomi B.
1984 "Women and politics: the real majority." In Jo Freeman (ed.) Women: A Feminist Perspective. Palo Alto, CA: Mayfield Publishing Co.

Mansbridge, Jane J.
1983 "The ERA and the gender gap in Illinois." Paper presented at the meetings of the Midwest Political Science Association, Chicago, IL.

Miller, Arthur H., and Oksana Malanchuk
1983 "The gender gap in the 1982 elections." Paper presented at the 38th Annual Conference of AAPOR, Buck Hill Falls, PA.

National NOW Times
1980–81 "Women vote differently than men; feminist bloc emerges in 1980 elections." December/January, p. 1.

National Organization for Women
c.1981 Women can make the difference, 20 pp., n.d. (probably late 1981).

Opinion Outlook
1981 "Reagan and women." September 7, p. 8.

Perlez, Jane
1984 "Women, Power and Politics." The New York Times Magazine, June 24, pp. 23–73.

Poole, Keith T., and L. Harmon Zeigler
1982 "Gender and voting in the 1980 presidential election." Paper presented at the meetings of the American Political Science Association, Denver, CO.
forthcoming Women and Political Life. New York: Longmans.

Riddlough, Christine R.
1981 "Women, feminism and the 1980 elections." Socialist Review, March/April, pp. 37–44.

Smeal, Eleanor
1984 Why and How Women Will Elect the Next President. New York: Harper and Row.

Stanwick, Kathy A., and Katherine E. Kleeman
1983 Women Make a Difference. New Brunswick, NJ: Center for the American Woman and Politics, Eagleton Institute of Politics, Rutgers—The State University of New Jersey.
Traugott, Michael W.
1982 "Gender and politics in the eighties." Economic Outlook USA, 9 (Autumn 1982):88–91.
Van Gelder, Lindsy
1984 "Ellie Smeal: how to tell the good guys from the bad guys." Ms., March, p. 55.
Walsh, Joan
1983 "Gender gap: women up the anti." In These Times, November, pp. 8–9.
Witt, G. Evans
1981 "Guest outlook." Opinion Outlook, March 9, p. 8.

Gender Differences in Voting for Female Candidates: Evidence from the 1982 Election

JOHN F. ZIPP AND ERIC PLUTZER

THE EMERGENCE both of a gender gap in the last two elections and of the National Organization for Women as a potent political force has led to increased pressure on both the Democrats and Republicans to run women for major elective office. Indeed, as the nomination of Geraldine Ferraro attests, both major parties are discussing various strategies, including nominating women, to secure the votes of women. Despite this, there is relatively little empirical research which documents (1) the relationship between the voting behavior of men and women and the gender of political candidates, and (2) more generally, the performance of women as office-seekers. We attempt to begin to rectify this situation by focusing specifically on these two matters. Our results will have important implications both for partisan strategies to deal with the gender gap and for the efficacy of women running for political office.

Previous Literature

During the last decade, the considerable attention focused on women and politics can be grouped into three broad areas: (1) how public

Abstract This paper analyzes voting in five 1982 elections in which women ran as major party candidates for high-level offices: Governor in Vermont and Iowa; and U.S. Senator in Missouri, New York, and New Jersey. Results indicate that the sex of the candidate generally has little impact on voting and that solid women candidates can attract cross-over votes, while weaker ones can lose them. The implications of these results are discussed in the concluding section.

John F. Zipp is an Assistant Professor and Eric Plutzer is a Ph.D. candidate, both in the Department of Sociology, Washington University, St. Louis, MO 63130, which supported, in part, this research. Data used come from surveys done by CBS, ABC and the St. Louis Globe-Democrat, KMOX television and radio stations in St. Louis. The authors wish to thank Betsy Bruce, Jack Flach, Dan Mittleman, Vince Narramore, Vicky Witte, Jan Yoder, and the Thursday on Friday Discussion Group for their help. The authors, each of whom contributed equally to this paper, are solely responsible for the analysis and conclusions.

Public Opinion Quarterly Vol. 49:179–197 © by the Trustees of Columbia University
Published by Elsevier Science Publishing Co., Inc. 0033-362X/85/0049-179/$2.50

policies affect women (e.g., Diamond, 1983; Gelb and Palley, 1982), (2) gender differences in political participation and attitudes (e.g., Baxter and Lansing, 1983; Goertzel, 1983), and (3) women's experience in public affairs, including barriers to the recruitment to and the holding of offices by political elites (e.g., Diamond, 1977; Kirkpatrick, 1974; Mandel, 1982). In examining how gender differences in voting affect the electibility of female candidates we are, in a loose sense, combining the emphases of the last two of these areas. We shall discuss each in turn.

SEX AND VOTING

Most of the literature on sex differences in participation has centered on party preference (Goertzel, 1983), voter turnout (Baxter and Lansing, 1983), and attitudinal differences (Goertzel, 1983; Spitze and Huber, 1980). *There is, however, no direct evidence from an actual election on the relationship between the sex of the voter and support for female candidates.* Obviously, a major impediment to this has been the relative paucity of women running for elective office. A number of studies have approached this issue indirectly. Ekstrand and Eckert (1981) gave various hypothetical male and female "candidate profiles" to a sample of 732 college students. Although in most cases there was no general bias toward or against female candidates, female voters favored a female candidate when that candidate was identified as "liberal leaning" but had no such inclination toward conservative females. Similarly, Hedlund, et al. (1979) in a survey of 233 registered voters found that women were more likely to say that they would be predisposed to vote for a hypothetical female candidate only in a certain circumstances: when she was not described as having small children and when there was no well-known opponent.

There is one other major line of indirect evidence. Beginning in 1937 and continuing to recently, Gallup polls have asked a version of the following basic question: "If your party nominated a woman for (president, Congress, etc.) would you vote for her if she were qualified for the job?" (Erskine, 1971). When asked about Congress, 10 percent of the men and 9 percent of the women surveyed in 1975 would not vote for the female candidate (Gallup, 1976; see also Erskine's 1971 review, and Welch and Sigelman, 1982).[1]

---

[1] Although Gallup did not ask about predispositions to vote against a male in a mixed gender race, when one of the preelection Iowa polls asked if Roxanne Conlin's gender affected the likelihood of voting for her, 14 percent of the respondents indicated that they would be less likely to vote for her, but 21 percent said that they would be more likely to do so (Des Moines Register, 1982a). In addition, the 1972 Virginia Slims Opinion Poll reported that 17 percent of the women and 7 percent of the men stated that they would be *more likely* to vote for a woman for president who was running against an equally qualified man, though a far greater number of both sexes indicated a greater likelihood of voting for the man in such an election (Harris, 1972).

These results not only suggest a serious handicap to women candidates (see below), but they also imply that there is little difference between the sexes in their support for female candidates. This contrasts sharply with the research noted above, leaving us in a position where it is not possible to draw any firm conclusions as to the effect that female candidates will have on the gender gap.

THE PERFORMANCE OF WOMEN AS CANDIDATES

There is ample documentation that women who attempt to enter the political elite, both in this country (Diamond, 1977; Mezey, 1980; Mandel, 1981; Welch, 1978) and others (Kornberg, et al., 1979) encounter serious difficulties in recruitment to candidacy and subsequent allocations of party resources. Given these and the Gallup poll results noted above, one might reasonably expect women candidates not to run as effectively as men.

Various authors using ecological analyses to address more specifically the performance of female candidates have reached contradictory results. For example, both Darcy and Schramm (1977) and Karnig and Walter (1976)found that women did not fare significantly worse than their male counterparts. Similar conclusions have been reached in analyses of parliamentary elections in Great Britain (Hills, 1981) and Australia (Mackerras, 1980). However, Ambrosius and Welch (1981), in a study of three midwest state legislatures, found that women polled significantly fewer votes and won significantly fewer elections than men.

Thus, we can draw contradictory expectations from the previous literature. If we relied on the Gallup and NORC poll results and the findings of Ambrosius and Welch (1981), we should expect women candidates to lose votes when running for state legislature, Congress, and the presidency. But according to other analyses, women do not perform significantly worse than men. Hence, we again are in a position where it is difficult to formulate solid conclusions.

Theoretical Expectations

SEX AND VOTING

On the basis of the literature cited above and the voting literature in general, we formulated four principal expectations regarding the relationship between sex and voting. First is the view that men and women will not differ in their likelihood of voting for a woman. Some support for this can be found in the recent Gallup (1976) and NORC polls (Welch and Sigelman, 1982) which showed basically no differences between men and women in their reported unwillingness to vote for a woman candidate.

Thus, this hypothesis suggests that sex will have no significant effect on voting for women candidates.

A second expectation argues that women are more likely to vote for a woman candidate solely on the basis of gender. This is consistent with Pomper's (1975) portrayal of a *dependent* voter: rather than relying on political factors, a dependent voter casts his/her ballot on the basis of social group membership. In this view, women—long afforded second-class political citizenship—will respond to the presence of a female candidate by voting for her. An extension or perhaps a confirmation of this view draws from the "group consciousness" explanation of black participation (Olsen, 1970): women who are more committed to women's issues should be especially likely to vote for the female contender.

Closely related to both of the above is a third possible expectation: the political leanings of the candidate will shape the relationship between sex and voting. The basis of the gender gap lies in issue voting: women are more likely to vote Democratic because the Democratic party has been more sympathetic to issues important to women, especially in terms of less military and more social spending (Goertzel, 1983). Ekstrand and Eckert (1981), although they did not use party identification, provided support for this with their finding that women were more likely to vote for a liberal-leaning female candidate. Based on this, our third hypothesis is that sex of the voter is unrelated to voting for a female candidate unless the candidate is more supportive than her opponent of issues of concern to women.

Our fourth hypothesis is also a further specification of hypothesis 2. This holds that sex has an impact on voting, but only for those without other strong cues on how to vote (i.e., Independents). The logic underlying this is that the propensity of women to vote for, and men not to vote for a female running for political office is evident only when other factors (the party identification of the voter) are not pertinent.

THE PERFORMANCE OF WOMEN CANDIDATES

We also have formulated a set of three expectations on the performance of woman candidates. On the basis of the Gallup and NORC poll results (Gallup, 1976; Welch and Sigelman, 1982) and the work of Ambrosius and Welch (1981), it can be expected that both men and women will defect from their normal vote in order not to vote for a woman (hypothesis 5). In distinction to this, if the ecological analyses of Darcy and Schramm (1976), Hills (1981), Karnig and Walter (1976), and Mackerras (1980) are correct, women candidates should not lose significant numbers of votes (hypothesis 6). Finally, it is possible that both are correct to a degree. Thus, we would expect that the gender of the candidate is unrelated to electoral performance. Similar to male candidates, some women will attract crossover votes and other will spur defections, due not to their gender but to their overall appeal as a potential officeholder (hypothesis 7).

Data and Methods

To test these hypotheses, we analyzed these relationships in five actual elections. All occurred in November 1982, and all were for major elective office. A short summary of the races is as follows:

IOWA GOVERNOR

Democrat Roxanne Conlin ran against Terry Branstad for the governorship held for 16 years by Republican Robert Ray. Branstad, former lieutenant governor, ran unopposed in the Republican primary, while Conlin, a former U.S. district attorney, beat two other candidates in the Democratic primary. Although neither candidate could be judged to be especially strong, Conlin led in the polls early in the campaign. During the campaign, however, it was disclosed that she had not paid any state taxes in 1981. This proved to be a major factor in her defeat, as 25 percent of the Branstad voters listed it as one of the reasons they voted for him (Des Moines Register, 1982b).

MISSOURI SENATE

Harriet Woods, a well-known state senator, opposed the incumbent, moderate Republican John Danforth. Woods, a liberal Democrat, was an upset winner in the hotly contested Democratic primary, receiving 43 percent of the vote against 10 male candidates. Initially a longshot, Woods—through a strong campaign—managed to draw even with Danforth in the pre-election polls the week prior to the election. A last-minute media blitz by Danforth was credited with preserving his narrow victory over her.

VERMONT GOVERNOR

The Vermont gubernatorial election matched incumbent Republican Richard Snelling against Democratic Lieutenant Governor Madeline Kunin. Kunin, who had served two terms as lieutenant governor, announced her candidacy after Snelling declared that he would not seek an unprecedented fourth term. Kunin, politically well known, was expected to become Vermont's first female governor until Snelling reentered the race. Snelling, a moderate Republican and a strong supporter of women's issues, won by a considerable margin (55 percent to 44 percent).

NEW YORK SENATE

This race pitted State Assembly Representative Florence Sullivan against Democratic incumbent Daniel Patrick Moynihan. Sullivan defeated two less conservative candidates in a Republican primary which attracted little interest. Moynihan's victory over Sullivan was the largest for a Senate seat in New York state history.

NEW JERSEY SENATE

One of the most watched races in 1982 had Republican iconoclast Millicent Fenwick running against Democrat Frank Lautenberg. A popular four-term U.S. representative, Fenwick was liberal on most social issues but supported Reagan's economic policies, the latter costing her the endorsement of NOW. Lautenberg, a businessman with very limited political experience, narrowly won a crowded Democratic primary and, with the support of labor, managed a close victory over Fenwick.

Fortuitously, not only did women run for these major offices, but two of the women were Republicans and three were Democrats. Also, NOW supported the three Democratic women and the two Democratic men. Furthermore, by focusing on contests for major elective office, we are fairly confident that voters would be relatively aware of the candidates, their positions, and their gender; further, the possibility that the vote for or against the female candidate represents the coattail effects of another vote are minimized.

It would have been ideal to be able to design a survey instrument and sampling scheme to be used for postelection polls in all five contests. Unfortunately, given when these candidates were nominated, it was not possible to do this. Therefore, we had to rely on secondary data. The best data are the exit polls done by the three major television networks. All do a national survey and also oversample certain states. And, although exit polls rarely have been utilized by social scientists, Levy (1983:66) concluded that they were "conducted with a degree of meticulousness which compares favorably to the highest standards of commercial and academic research."

Our original plan was to use all three exit polls in each state in order to strengthen the faith that could be placed in our results. However we were not able to do this for several reasons. First, since NBC would not release sample sizes, we did not use NBC's data in any of our analyses. Second, ABC oversampled only Iowa, New York, and New Jersey, and we used ABC's data in these three states. Finally, CBS oversampled all five states, but only in the New York and Vermont samples did they include enough information to test our concerns.[2] However, because CBS's estimate of the actual vote in New York was off by almost 10 percentage points, we felt that these data would not be suitable. Therefore, we were able to conduct the needed analyses through postelection interviews in Iowa (ABC), New Jersey (ABC), New York (ABC), and Vermont (CBS). A comparison of the actual proportion of the vote received by the female candidate to that estimated by these polls supports Levy's conclusion and gives us additional faith in them: Iowa: 46.9 percent (actual) vs. 45.7 percent (estimated); New York: 34.6 percent vs. 34.0 percent; New Jersey: 48.3 percent vs. 43.6 percent; Vermont: 44.4 percent vs. 48 percent.[3]

---

[2] They did not ask about voting in another election on that same day. We detail the importance of this below.

We were not as fortunate in Missouri. Since we could not obtain enough relevant information from the only exit poll (CBS) to oversample Missouri, we were forced to rely on a preelection survey sponsored by CBS-affiliate KMOX and the St. Louis Globe-Democrat newspaper. They employed a private research firm, Market Shares Corporation of Chicago, to conduct several preelection polls. The most extensive of these was a mid-October telephone survey of 456 likely voters several weeks prior to the election. The poll's estimate of the proportion of the vote received by Harriet Woods (49.5 percent) was extremely close to the vote she actually received (49.1 percent).[4] This gives us confidence in these data.

Our dependent variable in all the analyses that follow is a dichotomous one, with voting for the female candidate coded 1 and voting for the male coded 0. Because ordinary least squares regression is problematic with a dichotomous dependent variable, we have chosen to use logistic regression. A discussion of the problems of OLS and the benefits of logit analysis can be found in Dwyer (1983) and Hanushek and Jackson (1977).

To test the relationship between sex and voting for a woman, we regressed voting for the female candidate on a fairly extensive set of the independent variables typically used to explain voting (see the Appendix). First, we have the social and demographic factors of sex, age, race, place of residence (except in Vermont), and socioeconomic status, all of which have been found to be related to voting in previous studies (Milbrath and Goel, 1977). Second, in the ABC data, we have a series of questions tapping both perceptions of the performance of Ronald Reagan and various ideological issues: abortion, the nuclear freeze, and liberal-conservative self-placement. The Vermont data set also includes perceptions of Reagan and Reaganomics, along with liberal-conservative ideological self-placement. In Missouri, we have questions which focus directly on reasons for one's vote in the Danforth-Woods race, although these also tap issues pertinent to our analyses (e.g., position on abortion and the ERA). Finally, in each of the states we have a vote in another election: in Iowa, New York, and New Jersey, it is the election for Congress; in Vermont, it is for U.S. Senate; and in Missouri, it is for state auditor. We have included this voting variable so that we can examine the effects of a voter's sex net of a "baseline" vote. Thus, if sex is still significant after controlling for this it indicates the presence of a differential crossover vote between men and women.[5]

---

[3] Two points are pertinent here. First, these are percentages of the two-party vote. Second, ABC's lower estimate of the vote for Fenwick may be accounted for by the fact that they undersampled the elderly, one of Fenwick's bases of support (personal communication with David Scheer of Chilton Research, the firm which conducted the ABC exit polls.)

[4] It is worth noting that CBS's estimate of the Woods vote was 51.2 percent, a figure not as accurate as that of the preelection poll that we are using.

Because the most specific hypotheses on the sex and voting relationship are hypotheses three and four, we formulated our analyses to test them. (In so doing, all other hypotheses also will be examined.) To this end, we conducted separate logistic regressions for self-identified Democrats, Republicans, and Independents in each of the five states, giving us 15 logit models. If *hypothesis 4* is accurate, we should have sex related to voting for the female candidate only for Independents in all five states. For *hypothesis 3* to be correct, we should have sex related to voting for a female candidate only in Iowa, Vermont, and Missouri (where the candidate is more identified than her opponent with women's issues and supported by NOW), regardless of partisanship. It is easy to separate the candidates on their advocacy of women's issues in Iowa and Missouri, but not so in Vermont, as the incumbent Governor Richard Snelling was seen as being quite supportive of women's concerns. *Hypothesis 2* suggests that, net of all else, sex will affect voting in all five states and for those of all different partisan affiliations, and there will be a statistical interaction among sex, voting, and support for women's issues. Finally, *hypothesis 1* contends that sex will not be associated with voting under any of these situations. After these analyses are presented, we shall use the results to address hypotheses 5 through 7.

Before turning to our results, it is important to note the limitations of these analyses. In addition to our preference of employing the same postelection poll in all five states, it would have been better if the questions asked in the polls were more similar to ones asked in the standard SRC surveys. Beyond these, we are analyzing only five elections on one election day, and five very different ones at that. However, it is also true that if we obtain consistent results in these five different states, this should lend weight to our conclusions. Our conclusions, then, must be seen as somewhat preliminary, and our analyses need to be replicated and supported in future studies before any firm conclusions can be drawn.

---

[5] We are not implying that one's vote in another election affected how one voted in the election with the women candidates. Rather, we do contend that the social, demographic, and attitudinal variables, along with party identification, are imperfect indicators of a baseline or "normal" vote, and that including an actual vote improves our estimate of this baseline voting propensity and allows a more sensitive test of whether there is any differential crossover voting between men and women.

Results

The results of our logistic analysis of voting for a female candidate are contained in Tables 1–5, for Iowa, Missouri, Vermont, New York, and New Jersey, respectively. Turning first to Iowa, women who are self-identified Democrats or Republicans were no more likely than men to vote for Roxanne Conlin, but women who had an Independent partisan affiliation were more likely than Independent men to vote for her (.81; total effect). This effect essentially is unchanged by political attitudes and is net of the respondent's congressional vote on that same day.

We can highlight these results by computing the probability of voting for Conlin for men and for women in each of our three equations. In logistic regression independent variables do not have constant effects on

Table 1. Logistic Regression Results for 1982 Election, Iowa

	Democrats			*Republicans*			*Independents*		
Female	.16	.11	.01	−.07	−.12	−.13	.81*	.82*	.80*
Age	−.01	.08	.16	−.18	−.22	−.27	−.14*	−.12	−.12
Nonwhite	.67	.62	.24	−.10	−.42	.38	.62	.69	1.38
Small town	−.30	−.43	−.36	−.38	−.36	−.19	−.27	−.31	−.84*
Rural	−.60*	−.65*	−.90*	−.79*	−1.00*	−.93	−.42	−.42	−.60
Education		.44*	.37*		−.19	−.17		.10	.16
Income		−.19*	−.09		−.18	−.01		−.03	.15
Reagan: his job			1.15*			.18			.65
Reagan: economy			−.55			−.00			1.02*
Reagan: social security			.48			.72			.10
Reagan: inflation			.32			−.13			.20
Reagan: unemployment			−.40			−.20			.86*
Reagan: spending cuts			.21			1.90*			−.25
Abortion			.23			−.38			−.64*
Nuclear freeze			−.06			−.21			−.40
Liberal-conservative			−1.42*			−1.20			−1.03*
Congress vote			2.56*			1.08*			1.61
Constant	1.48	.54	[illegible].19	−.93	.87	−.93	.16	−.20	−2.94
Chi-square	8.65	24.20*	123.09*	7.01	12.26	72.54*	27.87*	29.14	214.04*
df	5	7	17	5	7	17	5	7	17
Sex × women's issues interactions									
Chi-square increment			.79			1.89			.93
df			1			1			1
Predicated probability of voting for the female candidate									
Men			.851			.067			.351
Women			.853			.059			.546

* $p < .05$

the dependent variable. Therefore, we need to evaluate the effects of sex both at its different levels (male and female) and at certain values of the other independent factors. Following convention, we have calculated the *probabilities of voting for Conlin* for men and women (in each equation) who are at the mean on all the remaining independent variables.[6] As can be seen at the bottom of Table 1, these probabilities indicate that there is very strong party-line voting for Democratic men (.851) and women (.853) and for Republican men (.067) and women (.059), with no differences between the sexes. However, Independents—who obviously could not vote party-line—differ sharply: the likelihood of men voting for Conlin was only .3[illegible], while it was .546 for women, over one and one-half times as much.

Very similar results were obtained in Missouri. Sex has no effect on voting for Harriet Woods for Democrats and Republicans, but it did have

Table 2. Logistic Regression Results for the 1982 Election, Missouri

	Democrats			*Republicans*			*Independents*		
Female	.58	.58	.51	.78	.80	.63	.59⁺	.61*	.46
Age	−.05	−.07	−.11	−.28	−.25	.10	−.04	.00	.01
Nonwhite	.80	.78	.37	.21	.33	.07	.66	.70	.63
Farm	−.61	−.59	−.38	.53	.57	1.36	.23	.20	.48
Income		−.07	.01		.14	.11		.20	.18
Vote: party id.			−.56			2.33*			.23
Vote: workers			−.82			.54			−.26
Vote: taxes			.33			.15			.11
Vote: unemployment			.33			.65			.54
Vote: opponent			−.24			.36			−.22
Vote: abortion			.02			−.42			.11
Vote: ERA			−.22			.23			−.08
Auditor vote			1.15			.54			.11
Constant	1.22	1.41	4.36	−2.34	−2.95	−12.70	−1.61	−2.15	−1.73
Chi-square	4.44	4.50	19.94	3.76	3.88	23.75*	−5.52	6.34	12.45
df	4	5	13	4	5	13	4	5	13
Sex × women's issues interactions									
Chi-square increment			3.14			2.06			3.85
df			2			2			2
Predicted probability of voting for the female candidate									
Men			.744			.065			.353
Women			.828			.116			.464

* $p < .05$
⁺ $p = .05$

---

[6] These calculations involved multiplying each independent variable's (except sex) mean by the corresponding logistic coefficient for that variable. These were then summed, along with the value of the constant, and termed XB. This was substituted into the following equation which yields probabilites (P): $P = 1/(1 + \text{exponent}(-XB))$. Because sex is coded so that males = 0 and females = 1, the above equation gives the probability that males voted for a female candidate. To obtain a similar result for female voters, the logistic coefficient for sex is added to XB, and the probability is then calculated in the same manner.

a significant impact for Independents (.59), with female Independents more likely than male Independents to have voted for Woods. The majority of this effect (80 percent) is direct, with the remainder mediated by political attitudes and a vote in the race for state auditor. These differences are illuminated with the probabilities shown at the bottom of Table 2. Although women Democrats (.828 vs. .744) and Republicans (.116 vs. .065) were more likely than their male counterparts to vote for Woods, neither of these differences were statistically significant and the magnitudes of the probabilities indicate strong voting along party lines. Female Independents (.464), however, were one and one-third more probable than men (.353) to have voted for Woods. The evidence from both Iowa and Missouri, then, gives strong support to hypotheses 3 and 4.

The results for Vermont, New York, and New Jersey stand in contrast to those for Iowa and Missouri. In Vermont, the voter's sex is not significantly related to voting for Madeline Kunin for Democrats, Republicans, and Independents. This is mainly due to party-line voting, as the probabilities that male and female Democrats voted for her were both over .80, while similar figures for male and female Republicans both were approximately .12, and for Independents just over .60. These results support hypothesis 1, which held that the sex of the voter is unrelated to voting for a female candidate.

In New York (Table 4), sex of the voter is unrelated to voting for Florence Sullivan for Democrats, Republicans, and Independents. This probably can be explained by two factors: (1) party-line voting, as indicated by the probabilities of both Democratic and Republican men and

Table 3. Logistic Regression Results for the 1982 Election, Vermont

	Democrats			*Republicans*			*Independents*		
Female	.29	.24	.19	.52	.51	.07	.18	.16	−.02
Age	−.18	.18	.13	−.14	−.17	−.07	−.28*	.28*	−.26*
Nonwhite	−.54	−.59	.45	.63	.46	.82	−.52	−.51	−.24
Professional, manager		−1.01	−.87		−.24	.26		−.61	−.74
Clerical, sales		−1.84	−1.01		−.23	.56		−.60	−.63
Blue collar		−1.27	−.73		.26	.77		−.70	−.86
Farmer		−1.13	−.16		−.62	.58		−.99	−.92
Income		−.26	−.24		−.29*	−.09		−.24*	−.14
Reagan: his job			1.85*			1.66*			1.62*
Reagan: economy			−.38			.13			.21
Liberal-conservative			−.68*			−.18			−.25
Senator vote			1.95*			1.47*			1.10*
Constant	.71	2.60	−.45	−1.53	−.57	−3.94	.94	2.18	−.84
Chi-square	2.75	10.74	69.08*	4.38	12.51	78.96*	7.23	16.42*	103.68*
df	3	8	12	3	8	12	3	8	12
Predicted probability of voting for the female candidate									
Men			.819			.121			.612
Women			.845			.128			.608

* $p < .05$

Table 4. Logistic Regression Results for 1982 Election, New York

	Democrats			*Republicans*			*Independents*		
Female	.00	−.06	.21	.18	.17	.40	−.33	−.33	.12
Age	−.13	−.09	−.14	.02	.01	.05	.13	.12	.10
Nonwhite	−.30	−.31	−.14	.45	.42	.16	−.86	−.83	−.24
Small town	.49	.63	.50	−.10	−.11	−.47	.58	.60*	.13
Rural	1.33*	1.47*	1.62*	.44	.42	.33	.96*	1.05*	.52
Education		.19	.23		−.03	−.02		−.14	−.06
Income		−.21*	−.25*		−.02	−.09		.12	−.03
Reagan: his job			−.45			−.56			−.83
Reagan: economy			−1.06			−.06			−.43
Reagan: social security			−.70			−.43			−.14
Reagan: inflation			.25			1.35*			−.15
Reagan: unemployment			.83			−.74			−.89*
Reagan: spending cuts			−.14			−1.02*			.18
Abortion			−.54			.27			.47
Nuclear freeze			.07			.42			.39
Liberal-conservative			.72			.88*			.70*
Congress vote			−2.14*			−1.28*			−1.35*
Constant	−1.75	−1.73	1.46	.37	.67	.59	−1.59	−1.60	.70
Chi-square	16.46*	21.84*	97.42*	5.15	5.42	60.94*	15.27*	18.33*	96.81*
df	5	7	17	5	7	17	5	7	17
Sex × women's issues interactions									
Chi-square increment			.09			.00			.10
df			1			1			1
Predicted probability of voting for the female candidate									
Men			.069			.646			.278
Women			.084			.731			.303

* $p < .05$

women voting for Sullivan (Table 4); and (2) the rather weak support she received among Independents. These results support hypothesis 3 most completely, but they also are consistent with hypothesis 1.

The New Jersey results are the least consistent with any of the hypotheses. They provide partial support for hypothesis 4, since sex was unrelated to voting for Millicent Fenwick, a Republican, for both Republican and Independent identifiers. However, among Democrats, women (.63) were more likely than men (.52) to vote for her. This latter result is consistent with hypothesis 2, which stipulated that women would be more likely to vote for women candidates.

The estimated probabilities for men and women in New Jersey clarify the above results. The first point that is apparent is further evidence of the strength of party ties: net of all else, both male and female Republicans were likely to vote for Fenwick 88 percent of the time, while the probability of voting for her among Democratic men was only .098 and among Democratic women .169. There is nothing in our data which allows us to determine why Fenwick did so relatively well among Democratic women. We are left to conclude that Fenwick's own characteristics (she was a

Table 5. Logistic Regression Results for 1982 Election, New Jersey

	Democrats			*Republicans*			*Independents*		
Female	.50*	.56*	.63*	−.30	−.18	−.05	.11	.12	.42
Age	−.14	−.15	−.09	.23*	.23*	.35*	.02	.03	.17
Nonwhite	−.75*	−.65	−.17	−1.37*	−1.20	−1.56	−1.00*	−1.06*	−.79
Small town	−.20	−.26	.26	−.28	− .24	−.37	−.03	.01	.06
Rural	.00	−.09	−.43		.38	−.12	.14	.22	.15
Education		−.06	−.02		−.06	−.02		.10	.27*
Income		.19*	.20*		.18	.22		.06	−.07
Reagan: his job			.61			1.01			−.31
Reagan: economy			−1.48*			−2.17*			−.51
Reagan: social security			−.34			−.76			−.02
Reagan: inflation			−.19			.06			−.53
Reagan: unemployment			.84			−.44			−.07
Reagan: spending cuts			−.92*			−.42			−.38
Abortion			.35			−.39			.14
Nuclear freeze			.03			.46			−.11
Liberal-conservative			−.75			.20			.76
Congress vote			−2.64*			−1.26*			−2.38*
Constant	−1.08	−1.72	2.61	.91	.21	3.17	.06	−.74	1.50
Chi-square	11.16*	16.87*	106.75*	10.22	12.81	60.87*	4.72	8.16	177.36*
df	5	7	17	5	7	17	5	7	17
Sex × women's issues interactions									
Chi-square increment			.49			.80			.30
df			1			1			1
predicted probability of voting for the female candidate									
Men			.098		.881				.524
Women			.169		.876				.626

* $p < .05$

liberal Republican who accepted no PAC funding and a past supporter of women's social issues) influenced these women to defect from their expected vote and vote for her.

Finally, one extension of our second hypothesis was that women who were more committed to women's issues would be more likely to vote for female candidates. Unfortunately, we have less than adequate measures of support for women's issues and we ask the reader to bear this limitation in mind. The closest that we could come was the respondent's feelings about abortion (in the Iowa, New York, and New Jersey samples) and whether the candidate's position on abortion and the ERA played a role in determining one's vote (in Missouri).[7] To test this notion, we created crossproduct terms between sex and these issues. We included these terms in each equation and compared the increments in chi-square associated with them. These increments, shown in Tables 1, 2, 4, and 5, indicate that there is no support for the idea that, net of other factors, women who are more committed to women's issues are more likely to vote for female candidates.

---

[7] There were no data in the Vermont survey which even allowed for these rough approximations.

In sum, we can conclude on balance that there is much more consistent support for our hypotheses 3 and 4 than for 1 and 2. More specifically, our findings can be summarized as the intersection of hypotheses 3 and 4: *The gender of voters is related to voting for female political candidates primarily among self-identified Independents voting in elections with women candidates who are closely identified with issues of importance to women*. The only case in which this was not upheld was Vermont, and that may very well be due to the fact that Kunin's male opponent also shared similar concerns. Despite this, from these findings we do not know if the sex gap exists because these women are more likely to vote for the woman or because men are less likely to do so. Determining this amounts to a test of hypotheses 5–7: do they tend to lose votes (hypothesis 5), not lose them (hypothesis 6), or like men, win some and lose some (hypothesis 7)?

To test this, we used a procedure similar to the concept of a "normal vote" (Converse, 1966). Basically, this contends that a normal vote is a product of party identification and short-term forces (e.g., issue positions). To construct our normal vote, we took advantage of this and other information we had concerning the respondent's vote in another electoral contest that same day. For Iowa, New York, and New Jersey, the election was for congressional representative; for Vermont, it was for U.S. Senate; for Missouri, it was for state auditor. We calculated an expected or typical vote by performing a logistic regression of the respondent's vote in this other election on the same factors used to explain voting for the female candidate. Then, as previously, we computed the probability that males and females voted for the candidate of the same party as the woman (e.g., voted Democratic for Congress in Iowa).[8] We shall use these probabili-

---

[8] We recognize the possibility that, if these other races had women candidates, our estimate of the normal vote would be biased. This was not a problem in Vermont and Missouri, since we used races with two male candidates as our baseline elections. Unfortunately, because we could not identify the congressional districts of voters, we could not assess this directly in the other three states. In New York, where four of the state's 34 congressional races had women as candidates, we replicated our analysis using vote in the governor's race and got highly similar findings. In New Jersey, where only one woman ran in 14 congressional races, the threat to validity is presumably quite small. Iowa presented more problems as women ran for Congress in two of its six districts. One was a Democrat, one a Republican, and both lost. It is only in Iowa that this factor could substantially affect our results, but even there it is likely to be minimal since in one the two races, the woman shared partisan identification with Roxanne Conlin. *If* defection occurred at the congressional level, it would tend to overestimate the normal vote in one district and underestimate it in the other. If these two districts were sampled equally (something we unfortunately cannot ascertain) these effects would cancel each other out.

Table 6. Predicted Probabilities of Voting for the Female Candidate and For a Candidate of the Same Party in Another Election[a]

	Democrats		*Republicans*		*Independents*	
	Female Candidate	*Party Vote*	*Female Candidate*	*Party Vote*	*Female Candidate*	*Party Vote*
Iowa						
Men	.851	.917	.067	.185	.351	.491
Women	.853	.884	.059	.167	.546	.526
Missouri						
Men	.744	.597	.065	.009	.353	.239
Women	.828	.613	.116	.004	.464	.114
Vermont						
Men	.819	.826	.121	.157	.612	.560
Women	.845	.814	.128	.176	.608	.595
New York						
Men	.069	.128	.646	.870	.278	.448
Women	.084	.101	.731	.813	.303	.536
New Jersey						
Men	.098	.077	.881	.765	.524	.324
Women	.169	.083	.876	.829	.626	.402

[a] In Iowa and Missouri, it is voting for the Democrat in Congress and the state auditor's race, respectively; in Vermont, it is voting for U.S. Senator; and in New York and New Jersey, it is voting for Republican for Congress.

ties as "typical" ones and compare the probabilities of voting for the female candidate to them.

These results are listed in Table 6. We can look at two types of comparisons: those in which sex of the voter was significant and those in which it was not. Turning first to those instances in which the voter's gender was related to voting for a woman candidate, we find contradictory evidence. In Iowa, a comparison of the estimated probabilities of voting for Conlin and Democratic for Congress indicates that the sex difference (among Independents) primarily was due to males deviating from their expected vote and not voting for Conlin. These men were equally likely to vote Democrat vs. Republican for Congress (.491), but over one-third as likely to vote for Conlin (.351).

The Missouri and New Jersey results differ markedly from these. In Missouri we are hampered by having to use a vote in the state auditor's race as our baseline, a race in which the incumbent Republican won handily. Given this limitation, which means that we likely are underestimating the normal Democratic vote, the findings reveal that both male and female Independents were more likely to vote for Harriet Woods than to vote for the Democratic contender in the auditor's race. The New Jersey findings are similar: among Democrats, both males and females were more likely to vote for Millicent Fenwick than to vote Republican for Congress. As in Missouri, women were much more likely than men to shift toward the woman candidate: male Democrats were approximately 25 percent more likely to vote for Fenwick than were typical Republicans, while women Democrats were over 100 percent more likely to do so.

A second set of comparisons looks at instances in which the sex of the voter was not statistically associated with voting for a female candidate. Essentially, in these (12 of the 15 possible instances) we have found that women and men defect toward or away from the female candidate in approximately equivalent proportions. But we have not yet established the direction of this defection. The findings reported in Table 6 do this, and again the results are mixed. In Iowa and New York, Conlin and Sullivan did worse than what might be expected. But, in Missouri and New Jersey, Woods and Fenwick gained votes over the normal vote for their parties. In Vermont, Kunin did better than what might be expected among male Independents and female Democrats, worse among Republicans, and about the same elsewhere. This would seem to support hypothesis 7: like their male counterparts, women candidates both gain and lose votes. And, much of it appears to depend on the candidate herself: where the candidate is a particularly solid one[9] (Woods or Fenwick), she can gain the votes of both men and women. Where she is a somewhat weaker one (Conlin or Sullivan), she can lose the votes of both men and women, and where both she and her opponent are strong (in Vermont), they each win some and lose some.

Discussion

We have two important findings of this research: (1) Sex is related to voting for a female candidate primarily among self-identified Independents in races in which the woman is identified as supporting issues which are important to women; and (2) strong female candidates can attract the crossover votes of both men and women, while weaker ones can lose the votes of men and women. Both of these findings have substantial implications.

At a scholarly level, they suggest that the votes of men and women are not a major reflection of the sex of the candidate. For the most part, voters seemed to ignore the candidate's sex and voted on the basis of more politically relevant factors. Given that the women candidates represented a fairly diverse political spectrum it is somewhat reassuring that sex was not a major factor in voting.

---

[9] Obviously, to measure the strength of a candidate by the vote received is tautological. We recognize this and we have attempted to characterize them on the basis of their qualifications and likely appeal, prior to the actual election. Thus, although judgments about the strength of candidates often derive from judgments about the candidates' likelihood of attracting votes, these conclusions are reached before the voting is done, they are typically made by party leaders if not by ordinary citizens, and they are not always right.

A second implication is more related to political practice. Our findings indicated that, in the large majority of possible instances (12 of 15), there was no major differential crossover vote toward or against female candidates by men and women. Moreover, we found that strong women candidates can attract the votes of men and women, while weaker ones can lose votes. This seems to imply that voters will treat female candidates the same way that they treat male candidates: if the candidate is perceived positively, citizens will crossover and vote for him/her. If not, citizens will crossover to vote against the candidate.

Finally, our findings speak to the practice of running female candidates for political office, with the Democrats doing so to maintain the allegiance of women voters and the Republicans doing so to reduce the gender gap. In either case, we would conclude that just running a woman is not likely to succeed unless that woman is a strong, serious candidate. Should either party choose to run women as tokens without using the same criteria as those for selecting male candidates, voters probably will respond by defecting from supporting that party.[10] However, should the Democrats and the Republicans, by ignoring the candidate's gender, nominate women whose backgrounds and credentials are conducive to attracting votes, citizens will vote for them. These findings support the position that the American voter is not swayed by the candidate's sex alone.

Although not directly comparable, it is useful to relate the implications of our analyses to the results of Geraldine Ferraro's candidacy. First and foremost, our results are supported as Ferraro had a minor impact on voting. *The New York Times* estimated that she gave the Democrats a net gain of eight-tenths of 1 percentage point (Dowd, 1984). As in our five statewide elections, the sex of the candidate appears far less important than other more politically relevant criteria. Secondly, we are struck by a major change as a result of the Ferraro candidacy. Prior to her nomination, there was a worry that women lose votes as candidates; after the election, the concern was that she did not *gain* votes. Perhaps both miss the point: female candidates are treated (by voters, at least) the same way as male ones, and we should not expect major shifts in either direction.

---

[10] At this writing, the Democratic Party's considerations of a woman for vice-president appear largely to be based on the symbolic value of having a woman on the ticket. Although our results are based on elections in which women were not chosen for this reason, one might infer from our analyses that—unless the woman chosen is perceived to be qualified for the position—there might be negative consequences for the Democrats. However, two points bear mentioning. It is common for some form of ticket-balancing, typically on regional or ideological grounds, to occur in presidential elections. It may be that gender will be regarded as another acceptable factor which goes into balancing a ticket, and thus carry no more risks than any other factor. The benefits/costs of this symbolic move should only be evident among those who alter their vote decision, in spite of party identification and issues, based on the vice-presidential nominees. The importance of the former in determining voting decisions clearly circumscribes the role that the vice-presidential nominee can play in shaping the vote.

Appendix: Data Employed

I. Data for Iowa, New York, and New Jersey are from ABC News/Washington Post Exit Polls. The data are archived by the Roper Association and the Inter-university Consortium for Political and Social Research, from which codebooks containing the exact question wording may be obtained.

II. The data for Vermont are from CBS News/New York Times Exit Polls and are also archived by ICPSR.

III. Data for Missouri are from a poll conducted for the St. Louis Globe Democrat Newspaper and KMOX-TV/KMOX-Radio. Below we give the exact question wording of the attitude items used as independent variables:

Everyone has their own reason to vote for a candidate. Which of these, if any, would be a major reason for you to vote for Danforth/Woods? *(ITEM READ)* Is this an *important* reason for you to vote for Danforth/Woods for US Senator . . . or *not*? (IF IMPORTANT, ASK:) Is this *very* important to you, or *somewhat* important? (Coding for each item: 1 = very important reason, 2 = somewhat important, 3 = not important.)

The items were read in the following order though the starting point in the list was rotated. The pneumonics employed in the tables are in parentheses. "Political party" (Vote: party id), "Cares about working people" (Vote: workers), "Opposed to tax increases" (Vote: taxes), "Would do more about unemployment problems" (Vote: unemployment), "Dislike other candidate" (Vote: opponent), "Candidate's position on abortion" (Vote: abortion), "Candidate's position on ERA" (VOTE: ERA).

IV. More complete information on the coding of independent variables from all three sources of data is available from the authors.

References

Ambrosius, Margery M., and Susan Welch
1981 "Women and politics at the grassroots: women candidates for state office in three states." Paper presented at the Annual Meeting of the Western Social Science Association, April 23–25, San Diego, CA.
Baxter, Sandra, and Marjorie Lansing
1983 Women and Politics: The Visible Majority, rev. ed. Ann Arbor: University of Michigan Press.
Converse, Philip
1966 "The concept of the normal vote." Pp. 9–39 in Angus Campbell, et al. Elections and the Public Order. New York: Wiley.
Darcy, R., and Sarah Slavin Schramm
1977 "When women run against men." Public Opinion Quarterly 41 (Spring):1–12.
Des Moines Register
1982a "Fierce, delicate battle for Democrats' hearts." May 23.
1982b "Democrats key to victory for Brandstad." November 21.
Diamond, Irene
1983 Families, Politics and Public Policies: A Feminist Dialogue on Women and the State. New York: Longman.
1977 Sex Roles in the State House. New Haven: Yale University Press.
Dowd, Maureen
1984 "The Ferraro factor little help at polls." Sarasota Herald Tribune, December 30, p. 1G.

Dwyer, James H.
1983 Statistical Models for the Social and Behavioral Sciences. New York: Oxford University Press.
Ekstrand, Laurie E., and William Eckert
1981 "The impact of candidate's sex on voter choice." Western Political Quarterly 34 (March):78–87.
Erskine, Hazel
1971 "The polls: women's role." Public Opinion Quarterly 35 (Summer):275–90.
Gallup, George
1976 The Gallup Opinion Poll Index. Report #128, March.
Gelb, Joyce, and Marian Lief Palley
1982 Women and Public Policies. Princeton, NJ: Princeton University Press.
Goertzel, Ted George
1983 "The gender gap: sex, family income, and political opinions in the early 1980s." Journal of Political and Military Sociology 11 (Fall):209–22.
Hanushek, Eric A., and John E. Jackson
1977 Statistical Methods for Social Scientists. New York: Academic Press.
Harris, Louis, and Associates
1972 Virginia Slims Public Opinion Poll.
Hedlund, Ronald D., Patricia K. Freeman, Keith E. Hamm, and Robert M. Stein
1979 "The electibility of women candidates: the effects of sex role stereotypes." Journal of Politics 41:513–24.
Hills, Jill
1981 "Candidates, the impact of gender." Parliamentary Affairs 34:221–28.
Karnig, Albert K., and B. Oliver Walter
1976 "Election of women to city councils." Social Science Quarterly 56 (March): 605–13.
Kirkpatrick, Jeane J.
1974 Political Women. New York: Basic Books.
Kornberg, Allan, Joel Smith, and Harold D. Clarke
1979 Citizen Politicians—Canada. Durham, N.C.: Carolina Academic Press.
Levy, Mark R.
1983 "The methodology and performance of election day polls." Public Opinion Quarterly 47:54–67.
Mackerras, Malcolm
1980 "Do women candidates lose votes? further evidence." Australian Quarterly 52 (Summer):450–55.
Mandel, Ruth B.
1981 In the Running: The New Woman Candidate. New Haven: Trichnor and Fields.
Mezey, Susan Gluck
1980 "The effects of sex on recruitment: Connecticut local offices." In Debra W. Stewart (ed.), Women in Local Politics. Metuchen, N.J.: Scarecrow Press.
Milbrath, Lester, and M. L. Goel
1977 Political Participation, 2nd ed. Chicago: Rand McNally.
Olsen, Marvin E.
1970 "Social and political participation of blacks." American Sociological Review 37:317–33.
Pomper, Gerald
1975 Voters' Choice. New York: Harper and Row.
Spitze, Glenna, and Joan Huber
1980 "Changing attitudes towards women's nonfamily roles: 1958–1980." Sociology of Work and Occupations 7 (August):317–36.
Welch, Susan
1978 "Recruitment of women to public office." Western Political Quarterly 31 (September):372–80.
Welch, Susan, and Lee Sigelman
1982 "Changes in public attitudes toward women in politics." Social Science Quarterly 62 (June):312–22.

Like Mother, Like Daughter: Intergenerational Transmission of DK Response Rates

RONALD B. RAPOPORT

A NAGGING problem in survey research is the "don't know" response. What makes "don't know" responses so difficult to deal with is that their interpretation is so ambiguous. As with other survey responses, the cause is partly the individual respondent and partly the question being asked. On the latter point, Schuman and Presser (1981) and Bishop et al. (1980) find a 20–25 percent increase in the DK response rate results when a filter, asking respondents, for example, if they have thought about the issue, is added to survey questions.[1]

---

[1] A more difficult problem, not dealt with here, is that the addition or subtraction of these "floating respondents" occasionally results in significant differences in both marginal distributions and in the interrelationships between two or more different attitudes (Schuman and Presser, 1981:126–141; Bishop, et al., 1980).

Abstract Attitude expression (and its opposite, DK response rate) is an important personal and political characteristic which is related to an individual's psychological and demographic traits. It is also transmitted from parent to child. In fact, it is transmitted to a greater degree than other political attitudes examined here, with the exception of partisanship. Owing to the fact that males show higher levels of attitude expression, and because this tendency is reinforced by cultural norms, the transmission process for attitude expression is conditioned by the sex of child and of parent. Compared with the father's impact, mothers are particularly influential in transmitting attitude expression to daughters. While fathers have a somewhat stronger effect on sons than do mothers, the difference is substantially smaller.

Ronald B. Rapoport is an Associate Professor of Government, College of William and Mary. The data for this study were made available by the Inter-University Consortium for Social and Political Research. Neither the Consortium nor the original collectors of the data are responsible for the analyses or interpretations presented here. The author would like to acknowledge the helpful comments and suggestions of Don Alexander, David Finifter, Fritz Gaenslen, Will Hausman, Eric Jensen, and the anonymous reviewers on an earlier version of this paper.

Public Opinion Quarterly Vol. 49:198–208 © by the Trustees of Columbia University
Published by Elsevier Science Publishing Co., Inc. 0033-362X/85/0049-198/$2.50

Although survey methods and question wording may retard or encourage DK responses, individual level variables show consistent relationships with DK response rates. The more educated the respondent, for example, the less likely to give DK responses. This finding is supported throughout the variety of question wording/filter experiments on issue questions, conducted by Schuman and Presser (1981) and Bishop et al. (1980).[2] It is reinforced, as well, by data from Gallup and Harris surveys. Using all Gallup questions between 1969 and 1971, and all 1970 Harris questions, Schuman and Presser (1981:145) found that only "14 percent of . . . 312 . . . items do not show a significant or noticeable DK-education association."

In addition to education, other demographic factors such as age, sex, race, and income also have consistent and significant effects on DK response rates (Francis and Busch, 1975). But it is not just demographic factors alone which relate to DK responses. Researchers find that psychological variables such as political involvement (Francis and Busch, 1975) and knowledge and generalized opinion strength (Schuman and Presser, 1981) are actually more important in predicting DK rates, when both they and demographic variables are used concurrently as predictors.

The consistency of these findings across questions of very different formats implies a strong individual level trait, that of opinionation (attitude expression). Further support for the importance of opinionation is provided by the facts that DK rates have been found to be extremely stable over time, in comparison with other political attitudes, and to relate quite strongly to other forms of political activity, such as trying to persuade others how to vote (Rapoport, 1982). In addition, it is almost unique among survey items in that it is a behavior that is observed directly in the survey situation.

Opinionation is not the same thing as attitude holding. One can hold an attitude without being willing to express it to friends or survey researchers. Similarly, we cannot distinguish between individuals with well-thought-out and stable attitudes and those who randomly pick a response category in order to appear knowledgeable. For this paper, opinionation is the more important variable. While attitude holding, alone, may affect one's own voting behavior or political participation, expressing attitudes reinforces and strengthens one's own attitudes, and can also influence the attitudes and behaviors of others.

But where does this orientation or trait come from? Most past research has focused on proximate determinants of DK response rates: education, age, political interest, etc. (see Converse, 1976–77; Francis and Busch,

---

[2] The single exception to this finding for both studies consisted of the made-up or unknown issues on which the more educated were more willing to admit to ignorance.

1975; and Ferber, 1966, for a sampling). Here, we look instead at the role of parental socialization. This makes sense since expressing attitudes is one of the very first political activities widely open to pre-adults, and a parental activity that a child cannot avoid observing. Is opinionation transmitted from parent to child, like some other attitudes, or is it transmitted through broader societal or peer socialization?

Parent-child transmission rates for traits and attitudes are extremely variable. In the only study including a U.S. national sample of youths (high school seniors) and their parents, Jennings and Niemi (1974) found a range in parent-child correlations (tau-b), from .47 for party identification, and .59 for 1964 presidential candidate support, down to .08 on attitudes toward big business and on allowing speeches against churches. The highest correlations were found on those attitudes which were most salient to the parents, most observable by the children, and concerning concrete rather than abstract political objects. Partisanship and candidate preference, for example, are easily and accurately perceived by the child, and frequently reinforced by the parent. They are cognitively simple attitudes, concerned with frequently encountered symbols. Similarly, the process through which voting turnout and other forms of activity are transmitted and reinforced is fairly clear, and the moderately high level of transmission (Tb's = .16 − .27) is understandable (Jennings and Niemi, 1981:88).

More general political orientations like political efficacy, cosmopolitanism, and political trust, however, show substantially lower levels of transmission (between .11 and .12), as reported by Jennings and Niemi (1974). This may be due to the greater integrative cognitive demands that transmission of these concepts puts on the child, both in terms of the diffuseness of parental cues and the complexity of these attitudes. The child must be able to recognize a variety of loosely related actions and statements as being related to one another and as comprising a unified attitude. But general orientations are important just because they are more abstract and less dependent on any particular issue.

In terms of transmission, we should not expect opinionation to be consciously reinforced in the same way as candidate preference or party identification, simply because it is difficult for the child to separate its form (i.e., to express or not to express attitudes in general) from its content (i. e., pro or con on specific attitudes). From a social learning perspective, we should expect that the child will have repeated exposure to a particular behavior of a valued model (i. e., parental opinionation or attitude uncertainty), and that this should produce imitation at reasonably high levels (Bandura, 1971; Cundy, 1979). We should expect, then, significant levels of transmission, but levels below those for particularly salient specific attitude objects, like party and candidate choice where not only modeling but purposive reinforcement is likely to occur, and where the attitude being transmitted is so clear.[3]

Method

Working with DK response rates is problematic. Differences in filters, question wording, and interview techniques can alter the number of DK responses (Schuman and Presser, 1981). To explain variance in DK rates, it is necessary to have a data set with individual variance in DK response rates. For example, the most widely analyzed data set which includes parents and children, that of Jennings and Niemi, shows such low numbers of DK responses, owing possibly to the agree-disagree question format, that there is little variance to explain.

The analysis here uses an eight-nation study directed by Max Kaase and Samuel Barnes. (A full explanation of the study, including sampling and index construction information, can be found in Barnes and Kaase, et al. (1979), and in the codebook to the data set distributed by ICPSR (#7777).) For this study, random samples of adults were interviewed in Great Britain, Germany, the U.S., The Netherlands, Finland, Switzerland, Italy, and Austria. The national surveys were carried out at various times between November 1973 and February 1976. In those homes in which an adult selected for the cross-sectional survey had a 16- to 20-year-old child who resided at home, that child was also interviewed, or if more than one was present, objective selection procedures were followed. In those cases in which a 16- to 20-year-old was a regularly selected member of the adult cross-section, an interview was also taken with a randomly selected parent, if present, in the household. This produced a total sample of 1635 child-parent pairs across the eight countries.

This is clearly not a random sample of 16- to 20-year-olds in these countries. Since the children still live at home, there might be closer ties between parents and children than would be true for a national sample of the same age group. This could result in higher levels of transmission of DK response rates, as well as most other attitudes, than would be true in a national sample of adolescents. However, by examining other parent-child attitude correlations in the sample, it is possible to infer, at least, the comparative degree of transmission of particular political orientations.

Although a multinational sample is not essential to test our hypotheses, it is useful. In Western societies with broadly similar social structures, similar rates of transmission should hold. Still, finding support for our hypotheses across different countries, with some variation in culture, certainly strengthens confidence in the findings (see Inglehart, 1981 on Western European political cultural differences in DK responses).

---

[3] Opinionation also differs from other orientations in that we should expect less feedback. Children's attempts to influence parents' attitudes relate to the direction of the attitude rather than to the expression/nonexpression dimension.

The questions over which the DK rate is computed include all of those with political relevance that were attitude questions rather than informational or fact questions. This left 116 questions for children and 109 for parents. (The sets of questions used were identical except for those asked of children about their parents.)

Findings

Table 1 indicates that intergenerational transmission of DK response rates does in fact occur. The correlation *(r)* across the entire sample, between parent and child DK response rate, is a robust .38. Even more interesting, this correlation is higher than that of many other general political orientations such as political efficacy, system responsiveness, and political interest. Only political trust shows as high a level of parent-child correlation, and only party preference shows a higher level.

It is, of course, the case that DK rates differ as a function of political

Table 1. Parent-Child Correlations on Political Orientations

Political Orientation	*Whole Sample Correlation*	*Standardized Country Correlations*
DK response rate[a]	.38 (1635)	.33 (1635)
Political efficacy[a]	.30 (1542)	.24 (1542)
System responsiveness[a]	.29 (1471)	.24 (1471)
Political trust[a]	.39 (1252)	.26 (1252)
Party preference	.58 (821)	.58 (821)
Strength of partisanship	.25 (1613)	.22 (1613)
Postmaterialism[a]	.16 (1543)	.14 (1543)
Political interest	.30 (1622)	.24 (1622)

[a] Dentoes multi-item scale.

Note: In this and the succeeding table, correlations are product moment correlations. *N*'s are in parentheses. For the standardized correlations, the country subsamples were standardized on each variable for children and parents separately, by subtracting the country mean on that variable (for either parent or child), and dividing by the standard deviation of the parent/child country sample. For each country, on each attitude, then, the mean for children and parents is 0, and the standard deviation is 1.

All scales are fairly standard except for *Postmaterialism*, and *System responsiveness*. *Postmaterialism* is based on questions asking respondents to rate certain values as to their personal importance. The preponderance of "postmaterialist" values (e. g., more popular participation, a greater emphasis on the roles of ideas vs. money) over "materialist" values (e. g., strong defense, economic growth) is the measure of *postmaterialism*. *System responsiveness* is a measure of how much respondents believe that political elites care about the ideas of, and keep in touch with people like themselves. Fuller descriptions of all scales can be found in Barnes and Kaase et al. (1979:537–86).

culture. Farah and Jennings (forthcoming), using the entire eight-nation adult sample ($N = 12{,}588$), found large differences across the eight countries in DK response rates, ranging from about 2 percent in West Germany to about 10 percent in Italy. Because internation variance is so great, computing a correlation across all eight countries has the effect of producing a higher correlation than those found within the individual countries. As a result, we recomputed parent-child correlations, standardizing both sets of responses by country so that, for both parents and children, on each attitude variable, the mean became zero, and the standard deviation one (this also corrects for differences in variances across different attitudes). Doing so actually strengthens our findings. As the "Standardized within Country Correlation" column in Table 1 shows, intergenerational transmission of DK response rates is higher (by at least .07) than for any other attitude, except for partisanship.

In addition, if we simply divide our sample (without standardizing variables) into country subsamples, we find further support for our hypothesis. (At the level of the individual country, DK rate transmission ranges from .20 in Finland to .55 in Italy.) In five of eight countries (data not shown), parent-child correlations on DK response rate are higher than any of the other general orientations, except for party, and in two of those (Netherlands and Italy) DK rate correlations surpass even partisanship. The three countries in which correlations for one or more general political orientations surpass those for DK rates (Great Britain, West Germany, and Finland) are those with particularly low DK rates (three of the four lowest), and therefore less variance to explain.

However, it is not impossible that the observed relationship could be at least partly spurious. If so, a strong candidate is education. It is likely that highly educated parents will, owing to their education level, have high levels of opinionation. In addition, the education environment in the home, even apart from the parents' levels of opinionation, may encourage both high educational aspirations in their children, and high levels of opinionation. In this case, controlling for parental education level should attentuate the relationship, and controlling for child's educational level should reduce it to insignificance. However, controlling for educational level of parent reduces the resulting partial correlation by only .01, and controlling for child's education has absolutely no effect on the parent-child correlation.

Our main hypothesis is thus confirmed. Not only are DK response rates reliable over time and useful in explaining levels of overt political activity among adults, as other studies have found, they are also transmitted from parent to child at a level equal to or greater than is true for other general political orientations, even strength of partisanship.

One of the most interesting findings about DK response rates is their correlation with sex. Consistently, in the U.S. and elsewhere (Holter, 1970; Bordieu, 1977; Cantril and Roll, 1971), females show significantly higher levels of DK response rates. The genesis of this sex difference can be found in a society's social norms. Not only do males dominate the political scene in all eight countries, but among mass publics, politics is viewed as more of a male endeavor (see, e. g., Levitt, 1967; Baxter and Lansing, 1980: esp. Tables 25 and 38–40; and Inglehart, 1981). In this study, for example, childen were asked how often they discussed politics, religion, sex, school, and work with each of their parents. Even though, in all eight countries in our sample, mothers are the more frequent parental discussion partner for the topics of school and work, religion, and sex, the father is the more frequent parental discussion partner on the topic of politics in seven of the eight countries. And in these seven the father is favored by more than an 18 percent advantage over the mother (i. e., the percentage discussing politics more with the father, minus the percentage discussing politics more with the mother.)[4] Even in the one country where fathers are not the more frequent partner, the U.S., the difference is small: 20.5 percent have more frequent political discussions with mother, versus 17.0 percent with fathers.

All of this implies a general and societally recognized male expertise advantage in political matters, at both the elite and mass levels. It is not so much, we would hypothesize, that females are discouraged from expressing attitudes, in any absolute way, particularly since we are looking at 16- to 20-year-olds in the mid-1970s. It is rather that males are encouraged to do so to a much greater degree.

In terms of family influence this leads to two expectations. First, since opinionation remains a sex-appropriate trait, particularly among the parental generation, females should be particularly influenced in DK response rates by their mothers. Where mothers are high in DK response rates, we should expect, of course, a particularly strong mother-daughter correspondence, based on identification and social learning theory, reinforced by social norms. But even when mothers show high levels of opinionation this, too, can be expected to produce corresponding levels in daughters, even without supportive social norms. This expectation is analogous to Kirkpatrick's (1974) finding that female political activists and officeholders are strongly influenced in their high level of activity by a politically active mother.

For sons, expectations are less clear. Opinionation is a more valued and expected behavior for males. Given the sex-linked character of opinionation, fathers' effects on opinionation of sons should be greater than that of mothers. This difference in parent effect should be less for sons than for daughters since it is possible for those parents of either sex who are high on opinionation to influence sons, particularly since other, nonfamily socialization agents are there to reinforce this parental influence. When parents are low in opinionation, we should expect nonparental agents to raise the levels for sons and to diminish the parent-child correspondence.

---

[4] In the areas of sex, work and study, and religion, mothers are the more favored parental discussion partner, across the eight countries, by an average of 29.0 percent, 12.1 percent, and 16.2 percent, respectively.

Looking first at parent-child correlations by sex (and now standardizing each variable separately for parent and child of each sex by country), we find that mothers do have a significantly greater impact on the DK rates of daughters than do fathers (.46 vs. .29; $p < .01$).[5] For sons, the father has a greater impact on the DK response rate of the child than does the mother (.39 vs. .29, $p < .06$).[6] Particularly interesting is the fact that this pattern is unique (as Table 2 shows) among general orientations we have examined. This is the only case in which we find both mothers having a significantly greater effect on daughters *and* fathers on sons. In fact, it is the only case in which, compared with fathers, mothers have a clearly greater effect on daughters, but not on sons. Furthermore, it is the only case showing a substantially greater difference in parent effects on daughters than on sons. What causes this uniqueness of opinionation is, as we have argued, its relationship to political sex-roles.[7]

---

[5] A difficulty in comparing mother-child with father-child correlations is that we are not comparing mothers and fathers of the same children. Because respondents were chosen randomly with respect to sex, and children in the house were likewise randomly selected, we have in essence, four unbiased subsamples of parent-child pairs (mother-daughter, father-son, mother-son, and father-daughter). Findings of differences among these subsamples can then be attributed only to differences in their sexual makeup. A more serious problem would exist if we were to try to control for variables like frequency of parental political discussion, likely to be related to political homogeneity of parents, which is, we know (Jennings and Niemi, 1974), strongly related to political transmission of most attitudes.

A second potential difficulty is that one parent may have a strong impact on the opinionation level of the child, not directly, but through the parent's effect on spouse's opinionation level, which then directly influences the child. If so, then using zero-order correlations between parent-child pairs to compare mother versus father effects could be misleading. This problem would only be significant, however, if there were a high correlation between opinionation of mother and father. Since the Barnes-Kaase study did not survey both parents of a given child, we cannot ascertain husband-wife correspondence on opinionation, or on any other attitudes. However, in Jennings' 1965 survey of a national sample of high school seniors and their parents (see Jennings and Niemi, 1974), both mothers and fathers of a random subset of the sample were surveyed on a wide range of attitudes and orientations (unfortunately their questions and interview technique produced only a handful of DK responses on attitude questions). Although it is certainly true that certain attitudes do show high levels of husband wife correspondence, conceptually, opinionation is a political orientation or trait that differs from these. Unlike partisanship and presidential vote, where tau-b correlations reach .60 and .82 respectively (Niemi, 1974:73), opinionation is not a likely subject for husband-wife discussion (i. e., separate from the direction of a particular opinion), and the dynamics of influence are not obvious. Even the sign of the relationship is not clear, since one highly opinionated partner may thwart the growth of opinionation in the spouse, as easily as encourage it. But even a general orientation like political efficacy, which is more likely to be discussed in the family, shows a gamma correlation between husband and wife of only .28 (Jennings and Niemi, 1971:72), and political interest is only correlated (tau-b) .27. For opinionation, we would expect the relationship to be smaller, but even these levels of correlation would not disturb our findings even if, as is very unlikely, they resulted totally from husband's direct influence on wife or from wife's on husband.

[6] We also used a *t*-test to test for interaction in the regression model. First for daughters and then for sons, child's DK rate was regressed on parent DK rate, sex of parent (a dummy

If our societal norm explanation is correct, then mothers and fathers should both have their greatest impact on children of the same sex when their level of opinionation is sex-appropriate, that is, when mothers are high in DK response rates, and when fathers are low in DK response rates. When we divide our sample into high and low DK rate mothers and high and low DK rate fathers, we find support for our explanation. Mothers have a greater impact on their daughters' DK response rates when they are lower in opinionation ($b = .58$) than when they are higher

Table 2. Parent-Child Correlations on Selected Orientations, Standardized by Country and Sex

Political Orientation	*Mother-Daughter*	*Father-Daughter*	*Mother-Son*	*Father-Son*
DK response rate	.46 (538)	.19 (280)	.29 (438)	.39 (378)
Political efficacy	.22 (501)	.29 (266)	.20 (416)	.29 (358)
System responsiveness	.28 (475)	.26 (252)	.13 (394)	.30 (349)
Political trust	.26 (389)	.24 (218)	.22 (347)	.34 (297)
Party preference	.63 (272)	.50 (138)	.61 (226)	.52 (185)
Strength of partisanship	.20 (527)	.21 (278)	.24 (434)	.21 (373)
Postmaterialism	.11 (508)	.15 (267)	.14 (413)	.16 (355)
Political interest	.26 (532)	.20 (278)	.26 (437)	.21 (374)

NOTE: Variables have been standardized within sex by country categories for both parents and for children.

---

variable), and an interaction term (sex times DK rate of parent). Using a t-test one can ascertain the significance of the sex by parent DK rate interaction. Doing so for daughters yields a coefficient for the interaction term which is positive and highly significant ($b = .35$; $p < .001$), indicating the significantly greater impact of mothers. Doing so for sons yields an interaction term coefficient which is negative, indicating the greater impact of males on sons ($b = -.08$), but which is significant only at the .10 level. (Normally, tests for interaction require F-test, but given a single interaction term, the F statistic is simply the square of the t statistic.)

[7] Partial correlations between parent and child DK response rates, controlling for parent education, had little effect on any of the four sets of parent-child pairs. In fact, in no case did the partial correlation decline from the zero-order level by more than .02. Controlling for child's education produced no decline at all, in the correlations, for any of the four sets of parent-child pairs.

(b = .47).[8] And correspondingly, fathers have a much greater impact on sons when they are high on opinionation (b = .52) than when they are lower (b = .30).[9]

The dominant role of males in the areas of politics makes sex differences in transmission patterns for attitude holding more similar to other behaviors, like running for office, than to attitudes like partisanship, cynicism, or even political efficacy, where sex-appropriateness is much less clear.

Conclusions

Opinionation is a significant aspect of political sex-role socialization. Previous research has shown that females, both adult and preadult, express fewer attitudes than males. What has not been clearly shown before is the role of parent-child transmission in perpetuating this sex difference. Here we find a strong effect. In fact, the level of opinionation transmission rates surpasses those for all other political attitudes examined here, except for partisanship. Most important is the sex-linked character of this parent-child transmission. For sons, fathers have a somewhat greater impact on their opinionation levels than do mothers, and this impact is greater at high levels of opinionation. On the other hand, the daughters' opinionation is much more strongly affected by that of mothers than of fathers, and the impact of the mother is greater when she has low levels of opinionation.

---

[8] Regression coefficients are used here because variances in both independent and dependent variables differ so greatly between samples of daughters with high and low DK response rate mothers, and between samples of sons with high and low DK rate fathers. Each standardized country by sex category is divided, at its mean, into high and low DK response rate groups, and the country groups were recombined for low and high parental DK response rates. The closer the regression coefficient is to one, the stronger the impact of mother or father on the child. As far as parent-opposite sex child correspondence is concerned, both mothers and fathers have the greater effect when their level of opinionation is high. For mother-son correspondence, the regression coefficient is .37 when mother is high in attitude expression, and .10 when she is low; for father-daughter correspondence, the coefficients are .28 and .03 respectively. All coefficients, except those for mother-son and father-daughter, under conditions of low parental attitude expression, are significant at the .05 level.

[9] Using another, more common statistical technique reinforces this finding. Child DK rates were regressed on parent DK rate and the square of parent DK rate, for both father-son pairs and for mother-daughter pairs. A significant square term indicates that there is a lack of linearity in the relationship, i.e., that the impact of parent DK rate on that of the child differs according to the level of the independent variable. The sign of the square term indicates whether the impact is greater when parent DK rate is high or low. For sons the coefficient for the square term is negative ($b = -.05$) and highly significant ($p < .01$), indicating a greater effect at lower levels of DK response rates. For daughters the coefficient for the square term is positive (b = .03), indicating, as expected, a steeper curve for higher level of DK responses. The coefficient is just significant at the .10 level.

References

Bandura, Albert
1971 Social Learning Theory. Morristown, NJ: General Learning Press.
Barnes, Samuel, and Max Kaase, et al.
1979 Political Action. Beverly Hills, CA: Sage Publications.
Baxter, Sandra, and Marjorie Lansing
1980 Women and Politics. Ann Arbor, MI: University of Michigan Press.
Bishop, George, Robert Oldendick, and Alfred Tuchfarber
1980 "Experiments in filtering political opinions." Political Behavior 2:339–370.
Bordieu, P.
1977 "Questions de politique." Actes de la Recherche en Sciences Sociales 44: 55–89.
Cantril, Albert, and Charles Roll
1971 Hopes and Fears of the American People. New York: Universe Books.
Converse, Jean M.
1976–77 "Predicting no opinion in the polls." Publican Opinion Quarterly 40:515–530.
Cundy, Donald T.
1979 "Affect, cue giving, and political attitude formation: survey evidence in support of a social conditioning interpretation." Journal of Politics 41:75–105.
Farah, Barbara, and M. Kent Jennings
1981 "Gender and political participation in Western democracies." Unpublished manuscript, Institute of Social Research, University of Michigan.
Ferber, Robert
1966 "Item nonresponse in a consumer survey." Public Opinion Quarterly 30: 399–415.
Francis, Joe, and Lawrence Busch
1975 "What we don't know about 'I don't knows'." Public Opinion quarterly 39:207–218.
Holter, Harriet
1976 Sex Roles and Social Structure. Oslo: Universities Forlaget.
Inglehart, Margaret L.
1981 "Political interest in Western European women." Comparative Political Studies 14:299–336.
Jennings, M. Kent, and Richard Niemi
1971 "The division of political labor between mothers and fathers." American Political Science Review 65:69–82.
1974 Political Character of Adolescence. Princeton, NJ: Princeton University Press.
1981 Generations and Politics. Princeton, NJ: Princeton University Press.
Kirkpatrick, Jeanne
1974 Political Women. New York: Basic Books.
Levitt, Morris
1967 "The political role of American women." Journal of Human Relations 15:25–35.
Niemi, Richard
1974 How Family Members Perceive Each Other. Princeton, NJ: Princeton University Press.
Rapoport, Ronald B.
1982 "Sex differences in attitude expression: a generational explanation." Public Opinion Quarterly 46:86–96.
Schuman, Howard, and Stanley Presser
1981 Questions and Answers: Experiments in Questionnarie Wording. New York: Academic Press.

Expectation/Disillusion and Presidential Popularity: The Reagan Experience

LEE SIGELMAN AND KATHLEEN KNIGHT

PREVIOUSLY we used national survey data from the early stages of the Carter presidency to test a series of hypotheses drawn from Stimson's (1976; 1976–77) explanation of recurring declines in presidential popularity and from Presser and Converse's (1976–77) extension of Stimson's work (Sigelman and Knight, 1983). This note uses data on Carter's successor, Ronald Reagan, to retest Stimson's interpretation.

Stimson argued that presidential popularity declines when unrealistically high expectations about what the president will accomplish give way to disillusionment; the public's lack of sophistication, in combination with overwrought campaign rhetoric, causes initial expectations to be so high and is thus responsible for the downturn in presidential popularity. From Stimson's arguments we distilled four propositions: (1) Over time, the decline in presidential popularity should be accompanied by an erosion in public expectations about what the president will achieve. (2) At any point in time, expectations and approval of the president should be closely interrelated. (3) Early in the president's term, the more educated (who are presumed to be more politically sophisticated) should

Abstract The Sigelman-Knight analysis of the "expectation/disillusion" interpretation of recurring declines in presidential popularity is replicated with data on the Reagan presidency. There is a connection between Reagan's popularity and expectations about what he would be able to accomplish, but (just as was the case during the Carter presidency, and directly counter to predictions) it was the more highly educated whose expectations of presidential performance were initially highest. Moreover, as time passed the size of the education-based gap in expectations of Reagan increased over time. These findings leave the expectation/disillusion interpretation in considerable doubt.

Lee Sigelman is Professor and Chairman, and Kathleen Knight is Associate Professor in the Department of Political Science, University of Kentucky.

Public Opinion Quarterly Vol. 49:209–213 © by the Trustees of Columbia University
Published by Elsevier Science Publishing Co., Inc. 0033-362X/85/0049-209/$2.50

have more realistic, and therefore lower, expectations than the less educated do. (4) As time passes, this gap should narrow considerably as the less educated become more disillusioned.

Analysis of data gathered in nationwide surveys between 1977 and 1979 revealed strong support for the first two propositions, but not for the latter two. Across a wide range of policy areas, there was no consistent relationship one way or the other between level of education and confidence that President Carter would achieve his goals; moreover, the size of the education-based gap in expectations remained unchanged over time.

Here we replicate our earlier analysis, using survey data from Ronald Reagan's first 13 months in office. These data are from CBS News/New York Times national telephone surveys in January, April, June, and September 1981 and January 1982, after which expectations concerning President Reagan's performance were no longer ascertained. In all these surveys except the first (which was conducted shortly before the inauguration), respondents were asked, "Do you approve or disapprove of the way Ronald Reagan is handling his job as president?" They were also asked whether "Ronald Reagan will or will not be able" to accomplish a number of policy goals (listed in Table 1). Some of these questions were asked only in January 1981, while others were asked intermittently thereafter.

According to Table 1, "relative approval" of Reagan (the number approving of his performance as a percentage of the number expressing

Table 1. Presidential Popularity and Expectations Over Time

	Jan. 1981	*Apr. 1981*	*June 1981*	*Sept. 1981*	*Jan. 1982*
Approval of Reagan		80.9	72.1	62.0	56.5
Agreement that Reagan would be able to:					
Reduce unemployment	55.4	47.9			30.8
Keep prices from going up as fast as they have been recently		47.4 [a]	42.8	33.4	33.6
Reduce the size of the federal government	65.6	71.3	57.9		61.6
Cut federal taxes	53.1	54.5	50.8		
Balance the federal budget within four years	40.1	38.3	37.0	35.4	25.8
Balance the budget *and* cut income taxes	27.5				
Reduce federal influence on people's lives	58.0				
Clean up the welfare system	63.4				
Strengthen military defenses	93.2			81.0	80.3
See that the U.S. is respected by other nations	92.3	92.0	83.4		79.6
Negotiate a nuclear arms treaty	43.1				

[a] In the April 1981 survey, a randomly selected half of the sample were asked this version of this question, and the other half were asked whether President Reagan would be able to "keep prices from going up all the time."

opinions) started off extremely high but dropped sharply during the year. Expectations about what Reagan would accomplish as president were also very high at the outset, the outstanding exception being cynicism about his pledge to balance the budget *and* cut taxes. Moreover, in Reagan's case, as in Carter's, declining popularity went hand-in-hand with diminished expectations. Although certain expectations proved more resilient than others, all were lower in January 1982 than they had been in January or April 1981.

As Stimson's interpretation predicts, then, the 1981–82 period witnessed concurrent drops in presidential popularity and in expectations concerning presidential performance. Also consistent with Stimson's interpretation, Reagan tended to be most popular among those who had the highest expectations of his performance. That is, on every expectation item and at every point in time there was a significant and substantial correlation between support for President Reagan and confidence that he would be able to solve a particular problem. Thus the first two elements of the expectation/disillusion theory are borne out by the Reagan experience, just as they were in the Carter presidency.

However, the data in Table 2 provide scant support for the idea that high early expectations are the province of the uneducated. In only two instances were there appreciable correlations between expectations and level of education, and one of these ran directly counter to the third hypothesis: in January 1981, confidence that President Reagan would be able to strengthen the nation's military defenses ran higher among more, not less, educated people. The less educated *were* more likely to expect Reagan to succeed in balancing the budget. More generally, though, there was no appreciable relationship between level of education and expectations concerning Reagan's performance in office. Controlling for party identification caused some moderate correlations to emerge between education and early expectations of Reagan's performance, but many of these party-specific correlations were positive, contradicting the third hypothesis.

Finally, in the entire sample and in each party group there was no discernible tendency for the modest education-based differences in expectations that prevailed in January and April 1981 to fade away by January 1982. If anything, the opposite was the case, for in several instances nonexistent or small early differences grew into fairly sizable ones. Moreover, this growth was not at all of the type contemplated in the expectation/disillusion interpretation. Rather, several correlations that had been negligible and/or negative turned moderate and positive with the passage of time—not only failing to wither away but also indicating that higher expectations were increasingly found among the more *highly* educated. Data not shown in Table 2 indicate that the more highly

Table 2. Correlations Between Expectations and Level of Education, by Party Identification

Agreement that Reagan Would Be Able to:	*Subsample*	*Correlation with Level of Education*				
		Jan.	*Apr.*	*June*	*Sept.*	*Jan.*
Reduce unemployment	All	−.028	−.013			.142
	Rep.	−.012	−.016			.177
	Dem.	−.114	−.056			−.103
	Ind.	−.003	−.096			.247
Keep prices from going up as fast as they have been recently	All		.000	.278	.329	.238
	Rep.		.124	.134	.285	.384
	Dem.		−.049	.365	.286	.148
	Ind		−.044	.235	.303	.195
Reduce the size of the federal government	All	.062	.137	.166		.194
	Rep.	.162	.211	.204		.071
	Dem.	−.028	.079	.134		.179
	Ind.	−.003	.073	.060		.173
Cut federal taxes	All	.086	.102	.168		
	Rep.	.143	.213	.217		
	Dem.	.064	−.047	.076		
	Ind.	.020	.181	.115		
Balance the federal budget within 4 years	All	−.235	−.143	−.133	−.045	−.226
	Rep.	−.147	−.206	−.115	−.050	−.257
	Dem.	−.299	−.237	−.281	.028	−.220
	Ind.	−.332	−.125	−.155	−.163	−.379
Balance the budget *and* cut income taxes	All	−.143				
	Rep.	−.104				
	Dem.	−.271				
	Ind.	−.147				
Reduce federal influence on people's lives	All	.046				
	Rep.	.155				
	Dem.	−.009				
	Ind.	−.055				
Clean up the welfare system	All	−.105				
	Rep.	.026				
	Dem.	−.144				
	Ind.	−.201				
Stregthen military defenses	All	.313			.310	.368
	Rep.	.346			.411	.429
	Dem.	.126			.250	.337
	Ind.	.354			.267	.328
See that the U.S. is respected by other nations	All	−.009	.037	.170		
	Rep.	.304	.148	.226		.177
	Dem.	.039	.016	.138		−.110
	Ind.	−.196	−.049	.115		−.023
Negotiate a nuclear arms treaty	All	−.156				
	Rep.	−.197				
	Dem.	−.120				
	Ind.	−.172				

NOTE: The correlation coefficient is gamma. All tables from which these gammas were calculated were "equiweighted," based on the algorithm given by Bruner (1976) to correct for unequal marginal distributions on the independent variable. The first entry for each item is for the entire sample; the second is for Republicans only; the third, for Democrats only; and the fourth, for Independents only.

educated clung to their high initial expectations or even raised those expectations, over the same period that expectations were declining among the less educated. Accordingly, these data are not at all consistent with the fourth hypothesis.

All in all, the data analyzed here provide even less support for the expectation/disillusion theory than did our earlier analysis of the Carter years. Evidence based on two very different political contexts, then, strongly suggests that the expectation/disillusion theory is in need of rethinking. During the Carter years, the best evidence for "disillusionment" was found among the *better educated*, who maintained for a longer period their initially higher expectations that Carter would restore trust in government, and then, apparently giving up, dropped their expectations below those of the less educated. Since Reagan assumed office the better educated have consistently maintained higher expectations than the less educated, except in regard to the naive expectation that Reagan would balance the federal budget. Overall, this suggests that it is the more politically sophisticated, and presumably more attentive, segment of the American public that is most responsive to presidential promise and performance.

References

Bruner, Jere
1976 "What's the question to that answer? Measures and marginals in crosstabulation." American Journal of Political Science 20 (November):781–804.
Presser, Stanley, and Jean M. Converse
1976–1977 "On Stimson's interpretation of declines in presidential popularity." Public Opinion Quarterly 40:538–41.
Sigelman, Lee, and Kathleen Knight
1983 "Why does presidential popularity decline? A test of the expectation/disillusion theory." Public Opinion Quarterly 47:310–24.
Stimson, James A.
1976 "Public support for American presidents: a cyclical model." Public Opinion Quarterly 40:1–21.
1976–1977 "On disillusion with the expectation/disillusion theory: a rejoinder." Public Opinion Quarterly 40:541–43.

Racism, Rational Choice, and White Opposition to Racial Change: A Case Study of Busing

MCKEE J. MCCLENDON

OVER the last three decades there has been a substantial reduction in white support for de jure racial segregation of schools, housing, and jobs and for laws against interracial marriage (Taylor, et al., 1978; Condran, 1979; Burstein, 1979; Smith, 1981). Most whites now favor equal opportunity for blacks in these institutions. Consistent with their acceptance of the principle of equal opportunity, most whites are also now willing to accept at least some contact with blacks in their own schools, neighborhoods, and homes (Taylor, et al., 1978; Condran, 1979; Burstein, 1979; Smith, 1981).[1] These improvements in racial attitudes suggest that tradi-

---

[1] Two caveats to this liberalizing trend are in order (Taylor, et al., 1978; Condran, 1979). First, the rate of increase in racial tolerance appears to have slowed or stopped during the 1970s. Secondly, not all racial attitudes have changed equally; the belief, for example, that "blacks shouldn't push where they aren't wanted" has shown little or no decline. This attitude will be examined further in this paper.

Abstract Despite significant declines in traditional prejudice, white resistance to desegregation and affirmative action has led some to argue that a new symbolic racism has emerged as the principal impediment to racial change in America. Studies of opposition to busing, however, suggest a nonracial or rational choice explanation; whites oppose busing because it is believed to have too many costs and not enough benefits. Using data from a medium-sized midwestern city, this study estimates a multiple indicator model for the effects of traditional prejudice, symbolic racism, and rational choice on four types of opposition to busing—attitudes toward two-way busing, one-way busing, protest, and white flight. Traditional prejudice and symbolic racism are found to be partially independent dimensions whose effects on busing opposition are entirely mediated by certain expected costs of busing. Thus, the findings support both racism and rational choice explanations. There is, however, no support for the notion that symbolic racism is a more important source of opposition to racial change than is traditional prejudice.

McKee J. McClendon is Professor of Sociology at the University of Akron, Ohio. This paper is a revised version of a paper presented at the 1983 meetings of the American Sociological Association in Detroit, August 31–September 4. The author wishes to thank Fred P. Pestello and the anonymous reviewers for *POQ* for their helpful suggestions. The author also expresses gratitude to Elizabeth Mutran and Robert Nash Parker for their consultation on LISREL. All deficiencies are the sole responsibility of the author.

Public Opinion Quarterly Vol. 49:214–233 © by the Trustees of Columbia University
Published by Elsevier Science Publishing Co., Inc. 0033-362X/85/0049-214/$2.50

tional racial prejudice may no longer be an important factor in perpetuating racial inequality.

If the principle of equal opportunity now enjoys widespread white support, such does not appear to be the case for many concrete efforts to change the racial status quo. Antipoverty programs, welfare for blacks, black militance, affirmative action, and busing for school integration are opposed by many whites (McConahay and Hough, 1976; Lipset and Schneider, 1978; Kinder and Rhodebeck, 1982; Pettigrew, 1979; Kinder and Sears, 1981). One reason for white opposition to affirmative action is that many whites now believe that opportunities for blacks to get ahead have improved greatly, and that today blacks have an equal or better than equal chance to get ahead (Kluegel and Smith, 1982; 1983). Thus, many whites seem to feel that there is little discrimination and that blacks are benefiting from preferential treatment or reverse discrimination.

The existence of such beliefs and attitudes about blacks' opportunity and programs for racial change, coupled with the decline in traditional prejudice, has led some to argue that a new "symbolic racism" has emerged (McConahay and Hough, 1976; Kinder and Sears, 1981). As conceptualized by McConahay and Hough, symbolic racism consists of negative affect toward blacks and the belief that black demands and government efforts to help blacks are violating cherished values embodied in the Protestant Ethic. Since blacks are believed to have equal opportunity, affirmative action and busing are symbolic of illegitimate black demands. Instead of being based upon traditional racist doctrines of white superiority and de jure segregation, symbolic racism is a hostile sense of racial injustice which emerged when the civil rights movement began to attack de facto segregation in the North (McConahay and Hough, 1976). Since traditional prejudice appeared to be at a low level in the North at that time, it has been stated that symbolic racism became the major source of opposition to racial change (Kinder and Sears, 1981:416). This notion implies that either traditional prejudice now has little or no effect on opposition to change, or minimally, that symbolic racism has a stronger effect than traditional prejudice, particularly in the North.

The validity of this thesis, however, has not been adequately tested by previous research. Given that symbolic racism and traditional prejudice are both said to be expressions of an underlying negative affect toward blacks (McConahay and Hough, 1976), symbolic racism may overlap to a great extent with traditional prejudice (Kluegel and Smith, 1983). With the possible exception of Bobo (1983), however, there has been little or no research to determine the degree, if any, to which symbolic racism is empirically independent of traditional prejudice. Moreover, there has also been no test of the notion that symbolic racism has a stronger effect on opposition to racial change than does traditional prejudice. This paper will test the independence and relative effects of the two types of racism through an analysis of white opposition to busing, a prototype for other types of affirmative action.

It has been noted (Kluegel and Smith, 1983) that the literature on symbolic racism often seems to equate all opposition to affirmative action programs with racism; opposition to busing, for example, has been treated as a measure of symbolic racism (Kinder and Sears, 1981). There is, however, a nonracial explanation for busing opposition which must be examined.

In the controversy over busing, outspoken opponents have frequently charged that busing disrupts cherished neighborhood schools, is too expensive, is unsafe, is harmful to children's learning, and increases racial tension (Hennessey, 1976). Many of these concerns are held by the general public as well (U.S. Commission on Civil Rights, 1973). Therefore, it has been argued that whites, and many blacks as well, oppose busing not because of racism, but because it is believed to have too many costs and not enough benefits (Orfield, 1978). This argument, which appears to be supported by several studies discussed below, will be referred to as a rational choice explanation.

According to rational choice theory (Heath, 1976; Blalock and Wilken, 1979; Riker and Ordeshook, 1973) individuals make choices between action alternatives on the basis of the expected outcomes of the actions and the subjective utility attached to each expected outcome. The rational individual will choose the alternative with the greatest utility summed over all expected outcomes. Rational choice theory, however, has little to say about where these expectations come from. This is an extremely important question which must be addressed if we are to better understand how people make choices in general, and choices about busing in particular.

According to Blalock and Wilken (1979:78–120), ideologies are often used to simplify what may be a highly complex task of predicting the various outcomes of an action alternative. In their view, ideologies may provide ready-made causal models of what to expect from a highly complex situation, thereby allowing the person to make reasonably quick decisions. The use of ideology in decision making in situations where there is little or no experience, such as busing, appears particularly probable.

Traditional racial prejudice is one form of ideology which may be used to estimate the expected outcomes of programs such as busing. The doctrine of white superiority, or black inferiority, should lead the prejudiced person to expect that the increased contact with blacks brought about by busing will have mostly negative outcomes (costs) for whites. Thus, the prejudiced person should oppose busing because he or she expects it to have many costs and few rewards. In sum, it is hypothesized that traditional prejudice may be an important cause of the expected costs and rewards which in turn affect opposition to busing; in statistical terminology, prejudice may have an indirect effect. This means that the rational choice explanation is not necessarily an alternative or rival to the racism explanation. The prejudiced person is simply seen as making a rational choice in the sense that he or she is attempting to maximize expected utility.

While there are good reasons for predicting a causal connection between traditional prejudice and some of the expected outcomes of busing, there do not appear to be clear theoretical grounds for predicting whether or not symbolic racism will affect these expected outcomes. Therefore, no hypothesis will be made about whether symbolic racism should have an indirect effect on busing opposition.

An extensive body of research on the effects of prejudice and racism has produced conflicting findings and conclusions. These studies have varied considerably in the type and number of indicators of prejudice that they have used. It appears, however, that studies which have used traditional indicators of prejudice have found either nonsignificant effects or relatively small effects (Kelley, 1974; Weidman, 1975; Stockton and Wayman, 1976; Casterline, 1977; Erbe, 1977; Stinchcombe and Taylor, 1980; Armor, 1980; McClendon and Pestello, 1982; McConahay, 1982). Furthermore, several studies have found that the expected costs and rewards of busing have significant effects on attitudes toward busing, even with controls for traditional indicators of prejudice (Stinchcombe and Taylor, 1980; Armor, 1980; McClendon and Pestello, 1982). The fact that traditional prejudice has little or no direct effect on opposition to busing while the expected costs of busing have relatively large effects has led to the conclusion that busing opposition is largely nonracial in nature (Stinchcombe and Taylor, 1980; Armor, 1980; McClendon and Pestello, 1982). None of these studies, however, investigated the possibility that prejudice might have an indirect effect through the expected costs and rewards. Thus, the hypothesis advanced in this paper has not yet been tested.

However, studies which have used at least some indicators of racism which are derived from, or consistent with, the symbolic racism theory have found significant and relatively strong effects (McConahay, 1982; Allen and Sears, 1978; Sears, et al., 1979; Sears, et al., 1980; Raffel, 1980; Useem, 1980).[2] Thus, there may be grounds for concluding that the busing research literature is supportive of the notion that traditional prejudice has been replaced by symbolic racism as a significant factor in American race relations. None of the studies using indicators of symbolic racism, however, have included measures of the expected costs and rewards of busing. And, with the exception of McConahay (1982), they have not controlled for traditional prejudice in order to estimate the independent effect of symbolic racism. In fact, indicators of traditional racism have often been mixed together with indicators of symbolic racism to construct racism scales. Thus, the relative importance of traditional prejudice, symbolic racism, and nonracial causes of opposition to busing are still very much at issue. This study uses multiple indicators of each of these three factors, as well as of several dimensions of busing opposition, in an exploratory attempt to clarify the roles of racial and nonracial causes of busing opposition.

---

[2] Useem's (1980) indicators of relative deprivation, which he found to predict support for protest against busing in Boston, are similar to the indicators of symbolic racism used by the other studies.

The model which will be used to estimate the effects of racism and rational choice on whites' opposition to busing is shown in Figure 1. The model specifies that traditional prejudice, symbolic racism, and the expected costs and rewards of busing may have direct effects on four dimensions of busing opposition:[3] attitudes against busing both black and white children (two-way busing), attitude against busing black children (one-way busing), intentions to protest court-ordered two-way busing, and intentions to avoid court-ordered two-way busing by removing children from the public schools (white flight). Most studies of the causes of busing opposition have focused on attitudes toward two-way busing. Armor (1980), however, found that prejudice had positive effects on opposition to the one-way busing of blacks and on white flight orientations, while Useem (1980) found that two of this paper's indicators of symbolic racism had an effect on protest. The model also specifies that prejudice and racism may have direct effects on the expected costs of busing.

The Sample

The data for this study were collected by telephone interviews with members of randomly selected households within the city limits of Akron, Ohio during March and April of 1981. Since directory sampling was used instead of the more expensive method of random digit dialing, households with unlisted numbers were not included in the sample. Rich (1977)

---

[3] Although specifying the expected costs as causes of busing opposition is consistent with rational choice theory and previous research, it is possible that attitudes toward busing may also be a cause of the expected costs, which would indicate choice rationalization. This study does not attempt to disentangle these reciprocal influences, but future research should try to do so, possibly by specifying some nonracial instrumental variables as determinants of perceived costs.

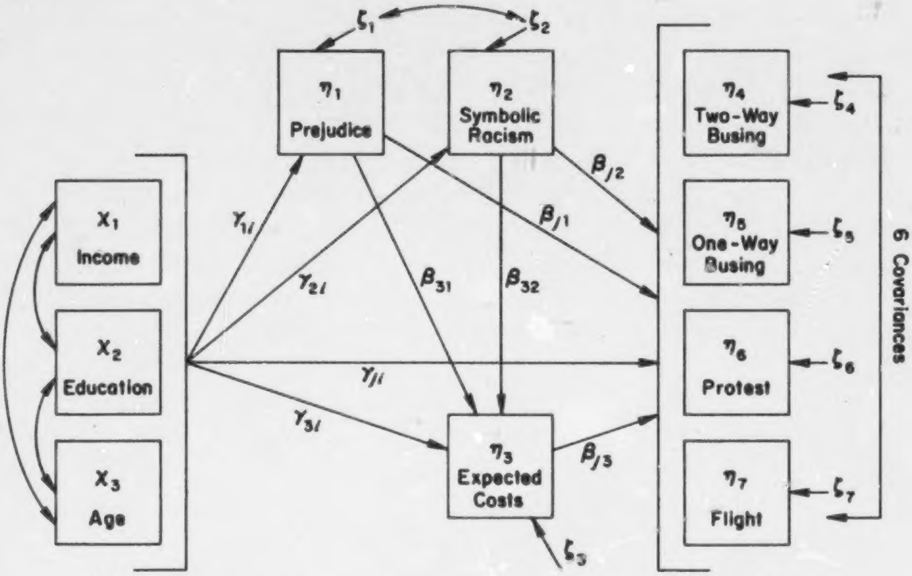

ζ_1
ζ_2
η_1 Prejudice
η_2 Symbolic Racism
η_4 Two-Way Busing
ζ_4
X_1 Income
X_2 Education
X_3 Age
γ_{1i}
γ_{2i}
β_{31}
β_{32}
β_{j2}
β_{j1}
η_5 One-Way Busing
ζ_5
6 Covariances
η_6 Protest
ζ_6
γ_{ji}
γ_{3i}
η_3 Expected Costs
β_{j3}
η_7 Flight
ζ_7
ζ_3

Figure 1. Structural Equation Model for White Opposition to Busing

has shown, however, that directory sampling produces little or no bias with respect to demographic variables.

For each selected household, the person whose name was listed in the directory was designated to be interviewed if that person was not married; if married, either the listed person or his or her spouse was randomly chosen to be interviewed. The 436 households that were called produced 242 completed interviews (55.5 percent), 18 partially completed interviews (4.1 percent), 140 refusals (32.1 percent), 3 no answers (0.7 percent), and 33 cases in which the selected person was unable to participate (7.6 percent). Of those interviewed, 63 percent are female, 17 percent are black, the mean age is 49.6, the median years of schooling is 12.2, the mean annual income is $18,605, and 38.7 percent have children under 18 years of age. In the 1980 Census, 55 percent of the adult population of Akron was female, 19 percent was black, the mean age was 44.7, the median years of schooling was 11.8, the mean income in 1979 was $17,631, and 32.2 percent of the households included children under 18 years of age (U.S. Bureau of the Census, 1983). Thus, the sample is somewhat older, somewhat more likely to have children at home, and more likely to be female than the adult population of Akron. A subsample of 190 whites is used in this study.

Akron is an appropriate setting for a study of busing opposition. First, as a nonsouthern city, it is a relevant location for testing the symbolic racism thesis which emphasizes the emergence and salience of this form of racism among northern whites. If the thesis cannot be supported in the North, where traditional prejudice is lowest, it is not likely to find support in the South. In terms of population (237,000 in 1980) and racial composition of the public schools (35 percent black at the time of the survey), Akron is rather typical of many large central cities in the North. Second, school desegregation has been a local issue—if not a volatile one—which has received periodic coverage in the news media. In April 1980 the U.S. District Court in Akron ruled that the Akron school officials and other defendants in a school desegregation trial were not guilty of illegal segregation. Furthermore, a small program of busing black students to predominantly white schools for ostensibly nonracial reasons exists in Akron.

Despite the fact that Akron is a rather representative city and a relevant location for the study of busing opposition, the small sample size, relatively low response rate, and the local sample itself dictate that the findings should be treated with the usual caution appropriate for exploratory case studies.

Observed Variables and Measurement Models

The measurement model for the endogenous variables (η_j) in Figure 1 is

$$y_i = \lambda_{ij}\eta_j + \epsilon_i,\ j = 1 - 7,$$

where the λ_{ij} are the unstandardized effects (factor scores) of the η_j on the observed y_i and the ϵ_i are random measurement errors. The indicators (y_i) for each η_j are shown in Table 1 along with their response distributions.

Questions one to five are traditional indicators of racial prejudice which have been frequently used to measure trends in attitudes toward integration (e.g., Taylor, et al., 1978; Condran, 1979; Burstein, 1979; Smith, 1981). Although DONTPUSH (y_4) has also been used as an indicator of symbolic racism (Allen and Sears, 1978; Kinder and Sears, 1981), a factor analysis (not reported) shows that DONTPUSH loads with the other traditional measures and not with the measures of symbolic racism. The LISREL analysis presented below also supports this conclusion.

Consistent with the general trend toward more favorable racial attitudes reported in national surveys, only a minority of Akron residents give prejudiced responses on the five traditional indicators. In addition to the five traditional items, WHITESCH (y_6) and WHITENGH (y_7) provide additional measures of preferences for racial segregation. The first seven items serve as the indicators for traditional prejudice (η_1).

Table 1. Percentage Distributions for Prejudice, Expected Costs-Rewards, and Busing Indicators ($N = 190$)

y_i	*Name*	*Question*	*Response*	%
		Racial prejudice (η_1)		
y_i	DINNER	How would you feel if a member of your family wanted to bring a black friend home to dinner?	Not object	81.6
			Object a little	10.0
			Object strongly	5.3
			DK/NA	3.2
y_2	MARRIAGE	Do you think there should be laws against marriages between blacks and whites?	No	71.1
			Yes	25.3
			DK/NA	3.7
y_3	KEEPOUT	White people have a right to keep blacks out of their neighborhoods and blacks should repect that right.	Disagree strongly	58.4
			Disagree slightly	25.3
			Agree slightly	6.8
			Agree strongly	6.3
			DK/NA	3.2
y_4	DONTPUSH	Some people feel that blacks should not push themselves where they are not wanted.	Disagree strongly	31.1
			Disagree slightly	27.4
			Agree slightly	14.2
			Agree strongly	16.3
			DK/NA	11.1
y_5	INBLOCK	How would you feel if a black family moved into your block?[a]	Would like it	4.7
			Not bothered	67.9
			Bothered a little	16.8
			Bothered a lot	7.9
			DK/NA	2.6
y_6	WHITESCH	Would you prefer to have a child of yours go to a school that is more than half black, mixed half and half, slightly more white than black, mostly white, or all white?	Doesn't matter/mixed	63.7
			Slightly more white	7.9
			Mostly white	12.6
			All white	10.5
			DK/NA	5.4
y_7	WHITENGH	Would you prefer to live in a neighborhood with more blacks than whites, that is mixed half and half, with a few more whites than blacks, that is mostly white, or all white?	Doesn't matter/mixed	33.2
			A few more whites	6.8
			Mostly white	39.5
			All white	17.9
			DK/NA	2.6
		Symbolic racism (η_2)		
y_8	UNFAIR	Now I would like to know if you feel that people are getting financially what they deserve—that is, what is fair? How fairly are black families being treated? Are they getting much more, somewhat more, somewhat less, much less, or about what they deserve?	Much less	5.8
			Somewhat less	19.5
			What they deserve	28.9
			Somewhat more	23.2
			Much more	3.2
			DK/NA	19.4
y_9	NOFAVORS	Irish, Italians, Jewish and many other minorities overcame prejudice and worked their way up. Blacks should do the same without any special favors.	Disagree	15.8
			Agree	77.4
			DK/NA	6.8
y_{10}	REVDIS	Many government programs designed to help blacks unfairly discriminate against whites.	Disagree	24.2
			Agree	63.7
			DK/NA	12.1

Table 1. ***(Continued)***

y_i	*Name*	*Question*	*Response*	*%*
		Expected costs-rewards (η_3)		
y_{11}	UPSET	Do you think that children who had to be bused would be very upset, somewhat upset, or not upset at all about leaving their schools?	Not upset Somewhat upset Very upset DK/NA	8.9 44.7 37.9 8.5
y_{12}	SAFETY	If a child of yours was to be bused to a mostly black school, would your child's safety worry you a lot, some—but not a lot, or not at all?	Not at all Some A lot DK/NA	14.2 34.7 45.8 5.3
y_{13}	HURTWHT	Would busing hurt the education of white children, improve their education, or make little or no difference?	Improve No difference Hurt DK/NA	6.3 61.1 24.2 8.4
y_{14}	TROUBLE	Would busing help black and white children to get along better, lead to trouble, or make no difference?	Get along No difference Trouble DK/NA	25.3 35.3 30.5 8.9
		Two-way busing (η_4)		
—		In general, would you favor or oppose the busing of black and white children from one school to another?	Favor Oppose DK/NA	4.7 91.6 3.7
—		In some cities the courts have found the school officials guilty of illegally segregating the schools. In such cases would you favor or oppose busing? [Not asked if previous response was Favor.]	Favor Oppose Not applicable DK/NA	17.9 65.3 4.7 12.1
—		If busing were only used as a last resort in such cases, would you favor or oppose it? [Not asked if previous response was Favor.]	Favor Oppose Not applicable DK/NA	17.4 53.7 22.6 6.3
—		Would you favor a law that would prohibit busing? [Not asked if previous response was Favor.]	Favor Oppose Not applicable DK/NA	46.8 10.0 41.1 2.1
y_{15}	TWOWAY	Two-way busing opposition scale	0 1 2 3 4	7.9 21.1 16.3 15.3 39.5
		One-way busing (η_5)		
y_{16}	ONEWAY	How would you feel if a considerable number of black children were bused to a school that a child of yours attended?	No objection Object somewhat Strongly object DK/NA	53.7 25.8 15.3 5.3
		If the courts ordered the busing of black and white children in Akron . . .		

(Continued)

Table 1. *(Continued)*

y_i	*Name*	*Question*	*Response*	*%*
		Protest intentions (η_6)		
y_{17}	MONEY	Would you give money to help fight the busing order in court?	No Yes DK/NA	44.7 43.7 11.6
y_{18}	LETTER	Would you write a lette against busing to either a newspaper editor or to your public officials?	No Yes DK/NA	33.2 62.6 4.2
y_{19}	MARCH	Would you participate in an organized protest march against busing?	No Yes DK/NA	75.3 20.5 4.2
y_{20}	BOYCOTT	Would you seriously consider keeping a child of yours at home to boycott the schools?	No Yes DK/NA	68.9 24.2 6.9
		White flight intentions (η_7)		
y_{21}	MOVE	Would you seriously consider moving to another school district outside of Akron to keep a child of yours out of the Akron schools?	No Yes DK/NA	69.5 25.3 5.3
y_{22}	TRANSFER	Would you seriously consider transferring a child of yours from the public schools to a private school?	No Yes DK/NA	50.0 45.8 4.2

[a] Those that said they would be bothered a little or bothered a lot were then asked, "If the black family had about the same income and education as you, would it bother you a lot, a little, or not at all?" Although the second question is the item that has been traditionally used, it was not used here since only about 8 percent of all respondents would be coded as prejudiced on this item.

Variables y_8-y_{10} are used as indicators of symbolic racism (η_2). UNFAIR is very similar to one of the measures used by McConahay (1982) and by Kinder and Sear (1981). NOFAVORS and REVDIS were modeled after two items constructed by Useem (1980), rather than after any particular items used in the symbolic racism literature. They have very good face validity, however, with respect to both the theoretical conceptualization of symbolic racism and a number of the specific indicators that have been used in that literature. Although only a minority of respondents believe that blacks are getting more financially than they deserve (UNFAIR), solid majorities feel that blacks should not receive any special help (NOFAVORS) and that whites often suffer from reverse discrimination (REVDIS).

Items 11–14 are indicators of some of the expected costs and rewards of busing (η_3). The items are similar to those used by the U.S. Commission in Civil Rights (1973), Armor (1980), and McClendon and Pestello (1982; 1983).

The next four questions in Table 1 were used to construct a scale of attitudes toward two-way busing (TWOWAY). Each successive item is assumed to indicate a greater intensity of busing opposition; thus, the items are assumed to form a Guttman scale. If a respondent favored busing on a lower intensity question, he or she was not asked the subsequent questions. A respondent's score on TWOWAY (y_{15}) is equal to the number of items on which he or she opposed busing. Attitude toward the one-way busing of blacks (η_5) is measured by a single variable, ONEWAY (y_{16}). Finally, there are four indicators ($y_{17}-y_{20}$) for protest intentions (η_6) and two indicators (y_{21} and y_{22}) for white flight intentions (η_7).

Findings

Maximum likelihood estimates for the coefficients of the structural equation model and the measurement model were obtained from the LISREL V computer program (Jöreskog and Söbon, 1981).[4] The chi-square goodness of fit test for the entire model shows that the model should be rejected ($\chi^2 = 287.21$, df = 232, $p = .008$). Jöreskog and Sörbon (1981), however, caution against using χ^2 to accept or reject a total model because, in most empirical work, the model is only an approximation to reality. The χ^2 is more appropriately used as just one indicator of the adequacy of fit of a single model and for testing the relative fit of different models.[5]

---

[4] Because of the small number of categories for most of the observed variables, the multivariate normality assumption of maximum likelihood is not met here. Although the maximum likelihood estimates may be used when the observed variables deviate moderately from normality (Jöreskog and Sörbon, 1981: I.29), more work needs to be done to determine if results are robust when poorly distributed variables are coupled with small samples.

The covariance matrix used to estimate the coefficients was calculated by pairwise deletion of missing values, and the sample size used in the LISREL solutions is 150. The measurement model for each unobserved variable with multiple indicators ($\eta_1-\eta_3$ and $\eta_6-\eta_7$) was estimated by setting $\zeta_i = 1$. The model fixes $\lambda_{ii} = 1$ for the single indicator constructs $\eta_4-\eta_5$. The measurement errors for these constructs (ϵ_{15} and ϵ_{16}) were fixed at $.1s_{15}^2$ and $.1s_{16}^2$, respectively, rather than at 0. Specifying a small amount of unreliability, although arbitrary, seems more reasonable than assuming perfect reliability. Furthermore, it does not make sense to estimate correlated measurement errors when ϵ is fixed at zero. Due to the large number of possible correlated errors, the .01 level of significance was used for estimating these correlations in order to reduce capitalization on chance. The modification indices were used to determine the order in which correlated errors were tested. Correlated disturbances in equations were estimated for the four busing dimensions since no causal relationships were specified between these factors. Only ζ_6 and ζ_7 were found to be correlated at the .05 level of significance. The five nonsignificant covariances were fixed at zero in the solutions to be presented.

[5] The goodness of fit index, which is independent of sample size, equals .667, and the index adjusted for degrees of freedom equals .543. The index measures the relative amount of variances and covariances accounted for by the model.

All of the observed variables have significant loadings ($p < .001$) on the constructs for which they serve as indicators (Table 2). DONTPUSH, which has sometimes been used as an indicator of symbolic racism, has the second highest loading on the traditional prejudice factor (η_1). The modification index (not shown) indicates that there would be virtually no improvement in the fit of the model if it were also allowed to load on the symbolic racism construct. Thus, it appears to be appropriate to treat DONTPUSH as an indicator of traditional prejudice.

The loadings for the measures of symbolic racism (η_2) and the expected costs of busing (η_3) are not as high, on the average, as the loadings for the indicators of the other constructs in the model. That is, the measurement reliability is lower for symbolic racism and costs than for the other factors.

Table 2. Standardized Measurement Coefficients for Model One

		$\lambda_{ij}{}^a$							
Variables		η_1	η_2	η_3	η_4	η_5	η_6	η_7	$s^2(\epsilon_i)$
y_1	DINNER	.47	—	—	—	—	—	—	.78
y_2	MARRIAGE	.64	—	—	—	—	—	—	.59
y_3	KEEPOUT	.73	—	—	—	—	—	—	.47
y_4	DONTPUSH	.70	—	—	—	—	—	—	.52
y_5	INBLOCK	.57	—	—	—	—	—	—	.68
y_6	WHITESCH	.61	—	—	—	—	—	—	.63
y_7	WHITENGH	.63	—	—	—	—	—	—	.61
y_8	UNFAIR	—	.42	—	—	—	—	—	.82
y_9	NOFAVORS	—	.52	—	—	—	—	—	.73
y_{10}	REVDIS	—	.40	—	—	—	—	—	.84
y_{11}	UPSET	—	—	.45	—	—	—	—	.80
y_{12}	SAFETY	—	—	.61	—	—	—	—	.63
y_{13}	HURTWHT	—	—	.42	—	—	—	—	.83
y_{14}	TROUBLE	—	—	.51	—	—	—	—	.74
y_{15}	TWOWAY	—	—	—	.95	—	—	—	$.10^b$
y_{16}	ONWAY	—	—	—	—	.95	—	—	$.10^b$
y_{17}	MONEY	—	—	—	—	—	.49	—	.76
y_{18}	LETTER	—	—	—	—	—	.45	—	.80
y_{19}	MARCH	—	—	—	—	—	.67	—	.55
y_{20}	BOYCOTT	—	—	—	—	—	.77	—	.41
y_{21}	MOVE	—	—	—	—	—	—	.77	.41
y_{22}	TRANSFER	—	—	—	—	—	—	.67	.55

Measurement Error Correlations[c]

	ϵ_3	ϵ_6	ϵ_{10}	ϵ_{15}	ϵ_{16}	ϵ_{17}
ϵ_4	.27	—	—	—	.53	—
ϵ_7	—	.27	.27	—	—	—
ϵ_{18}	—	.24	—	.63	—	.25
ϵ_{19}	—	—	—	—	.93	—

[a] $p < .001$ for all free coefficients.

[b] ϵ_{15} and ϵ_{16} were fixed at $.1s_{15}^2$ and $.1s_{16}^2$, respectively.

[c] $p < .01$ for all correlations.

In the structural equation model, the expected costs of busing have significant positive effects on each of the four types of busing opposition (Table 3, Model 1). Neither traditional prejudice nor symbolic racism, however, has a significant direct effect on any of the dimensions of busing opposition. All of these coefficients are small and/or in the wrong direction (negative). Although the estimated effects of traditional prejudice and symbolic racism on the expected costs are positive and moderately large, neither are significant at the .05 level, undoubtedly because the two racism dimensions are relatively highly correlated with each other ($r = .67$). Finally, education has significant negative effects on both prejudice and symbolic racism, and age has positive effects on both dimensions, although only the effect on prejudice is significant.

A number of additional structural models have been estimated (Table 4) in order to test various hypotheses about the independence of symbolic racism and traditional prejudice and their respective effects on opposition to busing. Model two constrains symbolic racism and prejudice to have equal effects on the expected costs of busing (i.e., $\beta_{31} = \beta_{32}$). The difference of χ^2 between models one and two provides a test of the null hypothesis that $\beta_{31} = \beta_{32}$. Since $\chi_2^2 - \chi_1^2$ is only .04 and is far from significant ($p > .80$), the null hypothesis cannot be rejected.

The tests associated with models 3–6 provide no support for unequal effects of prejudice and symbolic racism on any of the types of busing

Table 3. Standardized Coefficients for Structural Equation Models

	Dependent Variables						
Independent Variables	η_1 Prejudice	η_2 S.R.	η_3 Costs	η_4 TWOWAY	η_5 ONEWAY	η_6 Protest	η_7 Flight
Model 1							
x_1 INCOME	−.029	.090	−.002	.083	.002	−.072	.124
x_2 EDUC	−.256*	−.370*	.089	−.198	.048	−.142	.005
x_3 AGE	.236*	.134	−.059	.061	−.088	−.505*	−.380*
η_1 Trad. prej.	—	—	.380	−.063	.063	.032	.061
η_2 Symbol. racism	—	—	.483	−.061	.118	.006	−.305
η_3 Costs	—	—	—	.446*	.586*	.584*	.706*
R^2	.160	.148	.559	.193	.490	.562	.475
$r(\zeta_1\zeta_2)$		.623*					
$r(\zeta_6\zeta_7)$							.540*
Model 14							
x_1 INCOME	.003	.003	.025	.084	.011	−.071	.116
x_7 EDUC	−.289*	−.289*	.071	−.177	.014	−.148	.061
x_3 AGE	.209*	.209*	−.095	.046	−.071	−.498*	−.377*
η_1 Trad. prej.	—	—	.449*	—	—	—	—
η_2 Symbol. racism	—	—	.449*	—	—	—	—
η_2 Costs	—	—	—	.350*	.725*	.605*	528*
R^2	.152	.152	.588	.176	.513	.559	.413
$r(\zeta_1\zeta_2)$		.585					
$r(\zeta_6\zeta_7)$							.542

* $p < .05$

Table 4. Chi-Square Tests for Constraints on the Causes and Effects of Traditional Prejudice and symbolic Racism

		Diff. of χ^2 Test of Constraint						
M_i	*Constraint*	χ^2	*df*	*p*	$M_i - M_j$	$\Delta\chi^2$	Δdf	*p*
1	—	287.21	232	.008	—	—	—	—
2	$\beta_{31} = \beta_{32}$	287.25	233	.009	2−1	.04	1	$> .80$
3	$\beta_{41} = \beta_{42}$	287.21	233	.009	3−1	.00	1	$> .90$
4	$\beta_{51} = \beta_{52}$	287.24	233	.009	4−1	.03	1	$> .80$
5	$\beta_{61} = \beta_{62}$	287.22	233	.009	5−1	.01	1	$> .90$
6	$\beta_{71} = \beta_{72}$	287.57	233	.009	6−1	.36	1	$> .50$
7	$\beta_{j1} = \beta_{j2}, j = 3-7$	288.08	237	.013	7−1	.87	5	$> .95$
8	$\gamma_{11} = \gamma_{21}$	287.95	233	.008	8−1	.74	1	$> .30$
9	$\gamma_{12} = \gamma_{22}$	287.81	233	.008	9−1	.60	1	$> .30$
10	$\gamma_{13} = \gamma_{23}$	287.75	233	.008	10−1	.54	1	$> .30$
11	$\gamma_{1i} = \gamma_{2i}, i = 1-3$	289.19	235	.009	11−1	1.98	3	$> .50$
12	$\beta_{j1} = \beta_{j2}, j = 3-7$	290.22	240	.015	12−1	3.01	8	$> .70$
	$\gamma_{1i} = \gamma_{2i}, i = 1-3$							
13	$\beta_{j1} = \beta_{j2}, j = 3-7$	300.67	241	.005	13−12	10.45	1	$< .01$
	$\gamma_{1i} = \gamma_{2i}, i = 1-3$							
	$\sigma(\zeta_2\zeta_1) = 1$							
14	$\beta_{31} = \beta_{32}$	291.71	244	.020	14−12	1.49	4	$> .80$
	$\beta_{j1} = \beta_{j2} = 0, j = 4-7$							
	$\gamma_{1i} = \gamma_{2i}, i = 1-3$				14−1	4.50	12	$> .95$

opposition. Model 7, which pools all of the equality constraints of models 2−6, cannot be rejected, either. Thus, it appears that traditional prejudice and symbolic racism have equal, or nearly equal, effects on the expected costs of busing and on opposition to busing.

Turning to the measured causes of the two racism dimensions, the tests associated with models 8−11 indicate that the hypotheses that income, age, and education each have equal effects on prejudice and symbolic racism cannot be rejected. Model 12 specifies that both the measured causes and the effects of the two types of racism are equal. This set of constraints also cannot be rejected.

Model 13 constrains the covariance of the unmeasured causes of traditional prejudice and symbolic racism (ζ_1 and ζ_2) to be equal to one. Since these disturbances themselves are fixed at unity, this constraint has the effect of fixing $r(\zeta_1\zeta_2) = 1$. Furthermore, since model 13 also includes the constraints of model 12 (equal causes and effects), the model indirectly specifies that the dimensions of racism are perfectly correlated. Thus, model 13 assumes that symbolic racism and traditional prejudice are both expressions of a single underlying racism dimension and that there is only one racism factor affecting opposition to busing. Since the difference in χ^2 between models 12 and 13 is significant at the .01 level, the hypothesis that the disturbances are perfectly correlated is rejected.

Thus, it is concluded that symbolic racism and traditional prejudice are at least partially independent dimensions of racism.

Since none of the direct effects of prejudice and symbolic racism on busing opposition were significant in model 1, the final model (model 14) fixes these coefficients at zero, while retaining the constraints that they have equal effects on the expected costs and that they are equally affected by each of the exogenous variables. The tests associated with this model show that these constraints cannot be rejected. Thus, model 14 will be used as the best and most parsimonious model for the effects of racism on opposition to busing.

The bottom panel in Table 3 shows that education and age have significant negative and positive effects, respectively, on symbolic racism and traditional prejudice. Although their effects on traditional prejudice are well known, this study finds that they have the same effects on symbolic racism as well.

The correlation between the disturbances of the racism dimensions is .585; this correlation, plus their common dependence on education and age, produces a total correlation of .65 (not shown) between the two dimensions. Thus, although it has been found that symbolic racism and traditional prejudice are separate dimensions of racism, they are closely related. They also appear to affect opposition to busing equally and in the same way, indirectly; for both dimensions, racism increases the expected costs of busing, which in turn increases opposition to busing. Thus, the indirect effects that were hypothesized for traditional prejudice also occur for symbolic racism.

Since the effects of traditional prejudice and symbolic racism on busing opposition are entirely indirect in the final model, the relative size of their effects on the four dimensions of busing opposition is determined by the relative effects of the expected costs of busing. The expected costs have stronger effects on attitudes toward one-way busing and on protest and flight orientations than they do on attitudes toward two-way busing. Thus, prejudice and racism have their weakest effects on two-way busing attitudes, the dimension of opposition most frequently used in previous studies.[6]

---

[6] It may be of interest to note how the LISREL findings compare to those from an ordinary least squares (OLS) regression analysis. The regression findings are similar in that traditional prejudice and symbolic racism both significantly affect the expected costs of busing, which in turn significantly affect each of the four types of busing opposition. The regression results differ in that prejudice appears to have a stronger affect on the expected costs than does symbolic racism, and prejudice has significant direct effects on one-way busing attitudes and flight intentions. There are several differences between the LISREL and OLS procedures which undoubtedly contribute to these discrepancies. First, the multiple-item scales used in the OLS equations weight each item equally, whereas the LISREL measurement model allows

Discussion

The final model indicates that the expected costs of busing have a very strong effect on attitude toward one-way busing, strong effects on protest and flight intentions, and a moderate effect on attitude toward two-way busing. These effects support the rational choice explanation.

The final model also shows that these expected costs are strongly determined by traditional racial prejudice and symbolic racism, and that the expected costs appear to mediate entirely the effect of each dimension of racism. Thus, the final model also supports the racism explanation. The fact, however, that the effects of racism are entirely indirect indicates that rational choice and racism are not rival or different explanations for busing opposition.

The nature of the intervening expected costs help to explain the fact that they have a greater effect on attitude toward one-way busing than on attitude toward two-way busing. These costs consist mainly of the negative outcomes for white children which are expected to result from increased interracial contact in desegregated schools. For the one-way busing of black students, increased contact with black students is the only change affecting white students. White students do not have to ride buses, leave their neighborhoods, or attend black schools. Since the latter factors are relevant for two-way busing attitudes, but not for attitudes toward one-way busing, the expected costs of increased interracial contact have a stronger effect on opposition to one-way busing than on two-way busing.

Although the analysis of one-way busing attitudes helps to clarify the effects of racism and the expected costs of busing, the busing controversy has focused primarily on court-ordered two-way busing. The findings indicate that racism and the expected costs of busing have stronger effects on intentions to protest and flee court-ordered two-way busing than they do on attitudes toward two-way busing. Since there are costs to engaging in protest and flight (time, money, etc.), the most intensely opposed

---

them to have unequal loadings (i.e., unequal reliabilities). Second, LISREL corrects the estimated structural coefficients for attenuation due to measurement error, whereas OLS does not. Since the LISREL analysis (Table 2) indicates that the reliabilities of the measures for symbolic racism and the expected costs are lower than those for traditional prejudice, this would contribute to the greater relative importance of prejudice in the OLS analysis. Third, the LISREL analysis estimates several correlated measurement errors, whereas the OLS analysis assumes that there are none. And finally, LISREL simultaneously estimates each structural equation and the measurement model, whereas the OLS estimates for the structural equations are done one equation at a time. These differences should make the LISREL findings more valid than those of OLS. However, the two methods agree with respect to one important conclusion: symbolic racism is not a more important cause of busing opposition than traditional prejudice. The OLS estimates are given in Table A1.

persons should be the ones who are most likely to seriously consider protest and flight. This interpretation is supported by research showing that measures of attitude centrality and intensity are good predictors of "committed action," such as writing letters and contributing money (Schuman and Presser, 1981). Thus, these findings suggest that the intensity of two-way busing opposition is more strongly influenced by prejudice and racism than is the direction of such attitudes.

Finally, although this study has shown that symbolic racism and traditional racism are partially independent dimensions of racism, it has found no support for the notion that symbolic racism has replaced, or is more important than, traditional prejudice as a source of opposition to racial change. Both dimensions of racism appear to have approximately equal effects on busing opposition. Although the lack of support for the symbolic racism thesis in Akron raises questions about its validity, these findings need to be replicated with larger samples from other northern cities, or from the nation as a whole, before the thesis can be evaluated with a high degree of confidence.

Another important issue which needs to be investigated with a larger sample is whether the determinants of busing opposition for persons with children in the public schools are different from the determinants for other persons. McClendon and Pestello (1983), for example, argue that the effects of the expected costs of busing should be greater for parents of schoolchildren than for others. Although they found no support for this self-interest hypothesis, it would be desirable to test the model estimated in this paper for equality of parameters across the two populations. Such an analysis has not been conducted in this paper because of the very small sample of persons with children in the public schools ($N < 60$).[7] The issue, however, is an important one because the attitudes of these parents are likely to have a greater effect on policy makers than those of other persons.

---

[7] An OLS regression analysis with appropriate interaction terms in the equations found that prejudice had a significantly stronger effect on the expected costs of busing for parents with children under 18 years of age than for those without. Prejudice also had a significantly greater effect, and costs had a significantly smaller effect, on two-way busing attitudes and on protest orientations for those with children. Thus, prejudice appeared to be a more important factor and the expected costs a less important factor for those with school-age or preschool-age children. LISREL models were not estimated for the two populations because it is dangerous to use sample sizes smaller than 100 for such complex models (Long, 1983:59–60). Since discrepancies exist between the OLS and LISREL estimates for the entire population (see note 6), it would not be prudent to emphasize the OLS interaction effects.

Appendix

Table A1. Ordinary Least Square Regression Estimates of Standardized Coefficients[a]

Independent Variables	Dependent Variables						
	P	*SR*	*COSTS*	*TWOWAY*	*ONEWAY*	*PROTEST*	*FLIGHT*
INCOME	.008	.079	.020	.081	.017	−.040	.140
EDUC	−.234*	−.242*	.003	−.151	.036	−.150	.035
AGE	.203*	.031	−.011	.052	−.060	−.347*	−.250*
P			.365*	.053	.314*	.069	.175*
SR			.161*	.096	.114	.144	.027
COSTS				.198*	.354*	.283*	.232*
R^2	.114	.051	.194	.114	.356	.262	.192

[a] Scores for the scale P, SR, COSTS, PROTEST, and FLIGHT equal the means of the items with nonmissing data when at least half of the items do not have missing data, after recoding so that all items in a scale have the same range of values. Coefficients were estimated using pairwise deletion of missing data. The minimum pairwise $N = 160$.

*$p < .05$

References

Allen, H. M., and David O. Sears
1978 "Whites' opposition to busing in Los Angeles: is self-interest rejuvenated?" Paper presented at the annual meeting of the American Psychological Association, Toronto.

Armor, David J.
1980 "White flight and the future of school desegregation." Pp. 187−226 in Walter G. Stephan and Joe R. Feagin (eds.), School Desegregation: Past, Present and Future. New York: Plenum.

Blalock, Hubert M., and Paul H. Wilken
1979 Intergroup Processes: A Micro-Macro Perspective. New York: The Free Press.

Bobo, Lawrence
1983 "Whites' opposition to busing: symbolic racism or realistic group conflict?" Journal of Personality and Social Psychology (December) 45:1196−1210.

Burstein, Paul
1979 "Public opinion, demonstrations, and the passage of antidiscrimination legislation." Public Opinion Quarterly 43:157−72.

Casterline, John B.
1977 "Demographic correlates of attitudes towards busing and school desegregation: Detroit, 1976." Proceedings of the American Statistical Association: 448−53.

Condran, John G.
1979 "Changes in white attitudes toward blacks: 1963−1977." Public Opinion Quarterly 43:463−76.

Erbe, Brigitte M.
1977 "On the politics of school busing." Public Opinion Quarterly 41:113−17.

Heath, Anthony
1976 Rational Choice and Social Exchange. New York: Cambridge University Press.

Hennessey, Gary J.
1976 "An unfulfilled dream: a catalogue and analysis of the differing perceptions of busing: 1954−1976." Ph.D. dissertation Ohio State University.

Jöreskog, Karl G., and Dag Sörbon

1981 LISREL V: Analysis of Linear Structural Relationships by the Method of Maximum Likelihood. Chicago: International Educational Services.

Kelley, Jonathon

1974 "The politics of school busing." Public Opinion Quarterly 38:23–29.

Kinder, Donald R., and Laurie A. Rhodebeck

1982 "Continuities in support for racial equality, 1972 to 1976." Public Opinion Quarterly 46:192–215.

Kinder, Donald R., and David O. Sears

1981 "Prejudice and politics: symbolic racism versus racial threats to the good life." Journal of Personality and Social Psychology 40:414–31.

Kleugel, James R., and Eliot R. Smith

1982 "Whites' beliefs about blacks opportunity." American Sociological Review 47 (August):518–32.

1983 "Affirmative action attitudes: effects of self-interest, racial affect, and stratification beliefs on whites' views." Social Forces 61 (March):797–824.

Long, J. Scott

1983 Confirmatory Factor Analysis: A Preface to LISREL. Beverly Hills: Sage Publications.

Lipset, Seymour M., and William Schneider

1978 "The Bakke case: how would it be decided at the bar of public opinion." Public Opinion (March/April):38–44.

McClendon, McKee J., and Fred P. Pestello

1982 "White opposition: to busing or to desegregation?" Social Science Quarterly 63 (March):70–82.

1983 "Self-interest and public policy attitude formation: busing for school desegregation." Sociological Focus 16 (January):1–12.

McConahay, John B.

1982 "Self-interest versus racial attitudes as correlates of anti-busing attitudes in Louisville: is it the buses or the blacks?" The Journal of Politics 44:692–720.

McConahay, John B., and Joseph C. Hough, Jr.

1976 "Symbolic racism." Journal of Social Issues 32:23–45.

Orfield, Gary

1978 Must We Bus? Washington, D.C.: Brookings Institution.

Pettigrew, Thomas F.

1979 "Racial change and social policy." Annals of the American Academy of Political and Social Sciences 441:114–31.

Raffel, Jeffrey A.

1980 The Politics of School Desegregation: The Metropolitan Remedy in Delaware. Philadelphia: Temple University Press.

Rich, Clyde L.

1977 "Is random digit dialing really necessary?" Journal of Marketing Research 14 (August):300–5.

Schuman, Howard, and Stanley Presser

1981 Questions and Answers in Attitude Surveys. New York: Academic Press.

Riker, William H., and Peter C. Ordeshook

1973 An Introduction to Positive Political Theory. Englewood Cliffs, New Jersey: Prentice-Hall.

Sears, David O., Carl P. Hensler, and Leslie K. Speer

1979 "Whites' opposition to 'busing': self-interest or symbolic politics?" American Political Science Review 73 (June):369–84.

Sears, David O., Richard R. Law, Tom R. Tyler, and Harris M. Allen, Jr.

1980 "Self-interest vs. symbolic politics in policy attitudes and presidential voting." American Political Science Review 74:670–84.

Smith, A. Wade

1981 "Racial tolerance as a function of group position." American Sociological Review 46 (October):558–73.

Stinchcombe, Arthur L., and D. Garth Taylor
1980 "On democracy and integration." Pp. 157--86 in Walter G. Stephan and Joe R. Feagin (eds.), School Desegregation: Past, Present, and Future. New York: Plenum.
Stockton, Ronald R., and Francis W. Wayman
1976 "The busing issue and social change." The Michigan Academician 81 (Spring): 441–55.
Taylor, Garth D., Paul B. Sheatsley, and Andrew W. Greeley
1978 "Attitudes towards racial integration." Scientific American 238 (June):42–49.
U.S. Bureau of the Census
1983 1980 Census of Population and Housing. Akron, Ohio Standard Metropolitan Statistical Area, PHC 80-2-59. Washington, D.C.: U.S. Government Printing Office.
U.S. Commission on Civil Rights
1973 Public Knowledge and Busing Opposition: An Interpretation of a New National Survey. Washington, D.C.: U.S. Government Printing Office.
Useem, Bert
1980 "Solidarity model, breakdown model, and the Boston anti-busing movement." American Sociological Review 45 (June):357–69.
Weidman, John C.
1975 "Resistance of white adults to the busing of school children." Journal of Research and Development in Education 9 (Fall):123–29.

Face-to-Face Interviews and Mailed Questionnaires: The Net Difference in Response Rate

JOHN GOYDER

THE PROSPECT of a high response rate generally has been considered a decisive argument in favor of face-to-face interviews, over mailed questionnaires. A review of methodology textbook summaries of the literature on survey response rates reveals, however, a variety of estimates of the typical response likely from these two traditional forms of data collection. Many texts caution against expecting mailed questionnaire response from a general population to exceed about 30 percent (Labovitz and Hagedorn, 1971:52; Meyers and Grossen, 1974:177; Black and Champion, 1976:389). Babbie rates a 50 percent return as adequate (1973:165; also Bailey, 1978:136)—perhaps intending "adequate" as synonymous with "realistic"—and for several authors that indeed is essentially the upper limit to be expected from general populations (Orenstein and Phillips, 1978:229; Nachmias and Nachmias, 1976:107). For Kidder

Abstract Traditionally, as reflected in methodology textbook reviews of the survey nonresponse literature, interviews have been thought to achieve substantially higher response rates than mailed questionnaires. The paper describes a correlational design for assessing the typical response difference between these two forms of data collection, after controls for other factors known to affect survey response. Attention is given to change over time in the response difference between interviews and questionnaires. In the primary data set, the net response due to data collection amounts to some 7.5 percentage points, in the middle (30–70) range of survey response. The analysis quantifies the decline over time in interview response, and shows mailed questionnaires to be free from decline. A second data set reproduces the essential findings.

John Goyder is Associate Professor, Department of Sociology, University of Waterloo, Ontario, Canada N2L 3G1. The work reported herein was supported by a Canada Council Leave Fellowship held by the author during a stay at Cambridge University in 1978, and by a UW-SSHRC grant awarded in 1981. The author is greatly indebted to Cathie Marsh and others at the Social and Political Sciences Committee, Cambridge, for hospitality and research resources. Research Assistants were Norleen Heyzer and Kerry Davies at Cambridge and Jean Leiper and Olorunfe Taylor-Cole at Waterloo.

Public Opinion Quarterly Vol. 49:234–252 © by the Trustees of Columbia University
Published by Elsevier Science Publishing Co., Inc 0033-362X/85/0049-234/$2.50

(1981:150), the 50 percent ceiling holds true even allowing for follow-ups. Others, however, set the upper limit markedly higher. Fitzgerald and Cox (1975:53), for example, choose 60 percent, while, for True (1983: 218), "estimates of likely return rates range to 70 percent," a maximum also reported in Goode and Hatt (1952:173) and Cole (1980:62). Sanders and Pinhey (1983:152-53) cite 60 percent as an *average* and note that "some researchers have reported returns of 80 percent and more." Consistent 70 percent returns from general populations are not unrealistic, argue Weisberg and Bowen (1977:58), given energetic follow-ups (also Miller, 1977:79-81).

Estimated typical completion rate on interview surveys is 70–80 percent for Kidder (1981:150), 80–85 percent for Babbie (1973:171), 80–90 percent for Orenstein and Phillips (1978:229), over 90 percent according to Parten (1950:351), "about 95 percent" for Nachmias and Nachmias (1976:107), and "as much as 85 percent" for Sanders and Pinhey (1983: 154). Weisberg and Bowen (1977:36), introducing a factor to be emphasized herein, quote 90 percent interview response as "typical" during the 1950s, with rates "in the 80 percent range" common today. The greater response due to data collection by interview rather than by questionnaire could, from the above estimates, amount to anywhere from zero to some 60 percentage points.

In noting the variability of the textbook estimates, we do not deny that the accumulated wisdom on survey methodology presents a broad consensus that interview response is generally greater than mailed questionnaire returns, nor that each of the typical response figures cited by different authors accurately reflects at least some aspect of reality. The review of textbooks is effective, however, in emphasizing the variability in survey response according to both form of data collection and other factors. A "typical" survey response rate exists only insofar as there is a typical salience level of topic, sponsor, target population, and set of field procedures. In the unsystematic comparisons of questionnaire and interview response presented in methodology textbooks, these various factors have varied haphazardly.

The argument in this paper is that the survey methodology literature is now in a position to establish a net effect for form of data collection, controlling for other determinants of survey response by means of multivariate analysis. We develop this methodology herein, and establish the net margin between interviews and questionnaires, both as averaged over the history of survey research and as mediated over time.

Hypotheses

1. On average, face-to-face contact has stimulated higher net response rates than contact by mail.

2. The net response differential between interviews and questionnaires interacts with year of survey, converging over time.

Securing survey response is, for all but the most highly motivated of populations, largely a matter of overcoming resistance (Cannell and Kahn 1968:539), sometimes through "a long campaign of education" (Young, 1949:236). Thus, the social psychology behind hypothesis 1 is not that the public actively prefer personal contact in surveys so much as that they have found it more difficult to refuse a caller at the door than a solicitation arriving through the mail (Babbie, 1973:171; True, 1983:219; Sanders and Pinkey, 1983:154).[1] Such a view of nonresponse is not inconsistent with the exchange theory espoused in Gallegos (1974:3-9) and in Dillman (1978:12-19). For a population appropriately socialized into accepting the legitimacy of survey research, to refuse an interview is to incur the psycnic cost arising from the fieldworker's disapproval. Noncompliance costs less (though, as argued below, by a margin shifting over time) when contact is by mail and a questionnaire can plausibly be lost or forgotten, even if intentionally.

As to hypothesis 2, the study of the persuasive techniques leading to response maximization has cumulated effectively over the past several decades (Linsky, 1975). Researchers such as Dillman et al. (1974) have synthesized this knowledge into a coherent plan for raising response into the 70 percent range, effective even for long questionnaires and for general populations, and sufficiently infallible that "poor return rates can no more be excused than can inadequate theory or inappropriate statistics" (1974:755). Part of the increase in questionnaire response can be viewed within the present context as gross effect—as augmented "inputs" into a model of response—but the emergence of respectable mailed-survey completion rates probably has also been facilitated by the post-World War II decline in illiteracy, rated in the early 1950s as disqualifying at least 10 percent of the American population (Jahoda, et al., 1951:159). The interview fieldworker, unlike the mailed-questionnaire researcher, has always been seriously encumbered by failure to achieve contact, an increasingly troublesome problem, exacerbated by the deliberately unanswered doorbell (Ross, 1963). This, coupled with increasing refusal-after-contact, has eroded response rates on interview surveys during the past two decades (Steeh, 1981), enhancing the competitive position of questionnaires (Koenig, et al., 1977:433). Indeed, for some topics, a contemporary text has argued, "a mailed questionnaire may now elicit a higher response rate than a personal interview" (Nachmias and Nachmias, 1976:107).

---

[1] In a survey in Waterloo, Ontario, collected in 1982, we asked: "If you had to participate in a survey, but were offered the choice of having a questionnaire arrive through the mail, an interviewer call at the door, or a telephone interview, which would you prefer?" No method won a majority vote: 37 percent reported a preference for questionnaires, 45 percent interviews, and 11 percent telephone surveys (7 percent combination or no answer; cf., Sjoberg, 1955:423; Rogers, 1976:62; Groves, 1979:193). Recent literature that has influenced our view of nonresponse includes O'Neil (1979:229), DeMaio (1980:229–30), Stinchcombe, et al. (1981). A more extended statement of our conception of nonresponse appears in Goyder (1982a).

In the early days of survey research, even poorly controlled comparisons were probably not in error in suggesting a large, though unknown, net response advantage to interviews. Today, careful control is required before the magnitude, and even the direction, of the response differential according to form of data collection can be established. A correlational design for the attainment of this objective is described below.

Procedure

The literature on survey response has evolved from reports on single-survey case studies (e.g., Sletto, 1940; Kephart and Bressler, 1958)[2] to synthesizing review articles (e.g., Scott, 1961; Kanuk and Berenson, 1975; Linsky, 1975), to analyses in which case results from the accumulated literature themselves become unit of analysis in a quantitative analysis. Developed by Sudman and Bradburn (1974), working on the problem of "response effects," the technique was first adapted to the problem of nonresponse by Heberlein and Baumgartner (1978; 1981). By the late 1970s, the literature reporting experiments on survey response had reached sufficient proportions that Heberlein and Baumgartner could assemble some 214 "treatments" from 98 separate experiments dealing with mailed questionnaires (1978:448), a data base large enough to test for the simultaneous effect in a multiple regression equation of up to 18 determinants of response rate.

For present purposes, we have adapted the Heberlein-Baumgartner procedure to encompass the effect of form of data collection, compiling a sample of survey response rates for both interviews and questionnaires (along with limited data on other forms of enumeration, omitted here, due to paucity of observations). Our own data set draws from several sources. First, we relied, as did Heberlein and Baumgartner, upon the published survey methodology literature, though the bibliographic sources employed differed in the two compilations. Elsewhere (Goyder, 1982b:550-51), we have discussed the amount of overlap between the Heberlein-Baumgartner data and our own, noting some procedural differences between the two studies. The published literature reporting studies in survey response rates is largely from the U.S., with scattered examples pertaining to Canada or elsewhere. In order to achieve cross-national replication we augmented the data base for Canada, generating further examples of response rates by sampling survey-based articles from a Canadian periodical index, and contacting authors for field work information. Detail on this second sampling frame appears in Goyder (forthcoming).

---

[2] Case studies in the maximization of survey response continue, of course, to accumulate rapidly (e.g., Bradburn and Sudman, 1979:107-16; Singer, 1978).

Surveys were coded for response rate,[3] sponsorship (commercial, university, government, institution), type of population (general, employee, school or army, members of an organization, social categories), demographic features (such as homogeneity in education, occupation, sex, and urbanization), year of survey, and for field work procedures such as number of follow-ups or callbacks,[4] together with the use on questionnaires of precontacts, incentives, and special forms of final contact. Saliency of the survey topic was coded on a three-level classification—not salient, possibly salient, salient—devised by Heberlein and Baumgartner (1978:449). Coding specifications appear in their publication.

The coded variables reflect the information typically available from the published reports, and there is no claim to have measured every factor affecting response rates on surveys. In an ideal design, interviews would be coded for quality of field work indicators such as training procedures for interviewers, mean years of experience for interviewers, and expertise measures for principal investigators in surveys, and mailed questionnaire surveys might profitably be scored on persuasiveness of cover letter. Probably the most serious limitation to the data set is unavoidable inconsistency in the definition of nonresponse (see note 3); we suspect that the major portion of unexplained variance in the models to be presented below is attributable to this source of error. Thus, when relying upon published reports, it is impossible to construct the perfectly specified model of response. Other data sources—collecting information from the major survey houses, for example—undoubtedly would allow advances upon our own work, but would have their own limitations. The present data source is rich in capturing the historical legacy of surveying. Modern survey practice, reacting to the success of survey experts such as Dillman (1978), has probably evolved toward standardization on procedures. Our data set, in contrast, reflects a wide variety of survey practices. In Appendix 1, descriptive tabulations on the coded variables appear, allowing comparisons with other data sets on survey response, as they are compiled.

---

[3] Nonresponse is understood to include failure to contact, along with outright refusals. Ineligibles (e.g., nonexistent or vacant addresses) were, where possible, excluded from the denominator. Thus, we coded the nearest possible equivalent to what Shosteck and Fairweather (1979:211) term "final completion rate" (termed "response" by Kviz, 1977:265). In many surveys, those judged by fieldworkers to be unable to participate, on grounds of blindness, sickness, or illiteracy, were also excluded. "Nonresponse" can be defined both in more general and in more specific terms (Alwin, 1977:135,144). We have operationalized the concept as above mainly in conformity with the reports available from the data sources employed.

[4] Although coding interview response for each callback, we opted, for questionnaires, to record only final return rate after all follow-up attempts. The logic was that interview response rates on questionnaires are unstable unless the interval between waves is strictly standardized (see Vigderhous, 1977).

In the analysis reported below, surveys have been grouped by country—U.S. vs. other. Thus, only the U.S. cases from the sample generated from the methodology literature were selected, adding to the Canadian periodical index sample results from all non-U.S. experiments from the methodology literature. (The amount of European data was judged too small to sustain separate analysis, though we have addressed Canada-U.S. differences in Goyder, forthcoming). Table 1 records the sample sizes, classified by country and form of data collection. The analysis begins with the U.S. data, replicates using the non-U.S. sample, and then merges samples in order to achieve the broadest possible generalization about the consequences of data collection for survey response.

Previous work using the survey as unit of analysis in studies of nonresponse (Heberlein and Baumgartner, 1978; Eichner and Habermehl, 1981; Goyder, 1982b) has employed linear scoring of proportion responding. A consequent danger is some combination of values on independent variables in a model resulting in predictions exceeding 100 percent response. The previous researchers were not unaware of this liability (Heberlein and Baumgartner, 1978:458), but evidently concluded that the linear scoring allowed a simplicity and a face value interpretability that ensured dissemination of results to the maximum audience. For the present analysis, we have become persuaded in favor of a transformation. Thus, the dependent variable is an acrsine transformation (A = 2 arcsine

Table 1. Number of Observations in the Nonresponse Data Set, by Form of Data Collection and by Country

	Form of Data Collection				
Country	*Personal Interview*	*Mailed Questionnaire*	*Telephone*	*Drop-Off*	*Total*
U.S.	65	270	3	6	344
Canada	46	62	3	7	118
Britain	0	45	0	0	45
Other	1	9	0	0	10
Total	112	386	6	13	517

NOTE: Regression models in subsequent tables are based on these N's, and deal with personal interview and mailed questionnaire cases only.

$\sqrt{p}$) of the proportion responding in each survey. This tail-stretching transformation (Cohen and Cohen, 1975:256) in effect argues that near the 0 and 100 percent extremes the response rate on surveys loses elasticity.

Analysis

The study of survey response is a relatively mature field in the sense that a large body of literature, often reporting results from experimental designs, has removed much of the mystery from model specification. In the case of mailed questionnaire responses, Heberlein and Baumgartner's (1978) work, by achieving an empirical summary of a huge quantity of nonresponse literature, rendered the selection of independent variables particularly straightforward. The two key predictors of response, both in Heberlein and Baumgartner and in replications of their work, are number of follow-ups and salience of topic. The centrality of both factors is argued in other review articles on questionnaire response (e.g., Linsky, 1975), and so both were obvious main effects to be included in a general model of survey response. Special forms of follow-up contact, and the employment of cash incentives are also well-confirmed determinants of questionnaire response (Heberlein and Baumgartner, 1978; Goyder, 1982b), and clearly belonged in the general model. Dummy variables describing the sponsors of a survey and the type of population to be surveyed also hold proven consequences for response. From some five dummy categories describing type of population and another four coding sponsorship (Appendix 1), we settled upon two—general versus restricted population, and government sponsorship versus market, university, or institutional sponsorship—as adequately representing the importance of population and sponsorship.[5]

Number of questions has proved a significant but minor factor governing response to mailed questionnaires (Heberlein and Baumgartner, 1978:457; Goyder, 1982b:551). We elected to delete length from the model; the coding of length of interview schedule was hopelessly incomplete, leaving it impossible in any case to express the fact that length is probably more important for questionnaries than for interviews.

Finally, dummy variables for form of contact (face-to-face interview = 1, mailed questionnaire = 0) and year of study were required for tests of the hypotheses motivating the study. Year was entered as a dummy category distinguishing pre-1970 surveys from later ones. This scoring was selected after adjusting interview response (the mode of data collection where a temporal decline in response was expected) for the other factors in the model, and examining residuals for surveys grouped by decades, from 1930 onward. The residuals proved stable from the 1930s to the 1960s, showing a sharp breaking point between 1960 and 1970.

---

[5] In our own report on a replication of Heberlein and Baumgartner, the effect of general population appears as a minor effect, but the reference category in that computation does not include the high-response generating categories of employee and school or army populations. In the present analysis we opted for the simplification of merging all special populations, and contrasting the lower response obtained in general populations.

For parsimony, the analysis began by assuming just the existence of main effects in the general model, and introduced interactions only if persuaded by the data to do so. It was hardly to be expected that the *direction* of relationships between factors such as number of contacts and saliency would interact by form of collection. Both factors are known (e.g., Kish, 1965:550; Survey Research Center, 1976:8) to be crucial determinants of interview response. The issue was whether, for example, the consequences of an interview callback sufficiently resembled those of a questionnaire follow-up that both interviews and questionnaires could be coded on a single variable—number of contacts. Similarly, it seemed probable that the effect of population and sponsor would be germane to both interview and questionnaire response, and the modeling decision again concerned whether the strength of effect differed according to form of collection. Thus, interaction variables representing form of data collection by, successively, number of contacts, salience of topic, government sponsorship, and general population were tested. Attention was confined to the U.S. data, for initial model-building; along with being the larger data set, multicollinearity among predictors, a factor emphasized by Eichner and Habermehl (1981), in commenting on Heberlein and Baumgartner's work, is appreciably lower in the U.S. data (which includes some factorial designs in the compilation) than in the foreign sample. The effect of number of contacts, government sponsorship, and general population proved stronger on mailed questionnaires, probably reflecting the active role of interviewers at the doorstep in determining response, contrasted to questionnaires, which must "sell" themselves (Levine and Gordon, 1958:571) unassisted. But none of the interactions approached .05 statistical significance, and so a main effects model of response was adopted as starting point for analysis.

Parameter estimates for this main effects model, presented separately for the U.S. and for the foreign data sets, appear in Table 2. The multiple regression coefficients record the contribution increments on each independent variable make to A. A table of back-transformations (Cohen and Cohen, 1975:257) is needed to unravel the implications of the different determinants of survey response. Taking a largely arbitrary starting point, the intercept, at $A = 1.158$, back-transforms to a response of

Table 2. The Net Effect of Form of Data Collection, as Estimated in Two Data Sets (Dependent Variable: Arcsine Transformation, Proportion Responding)

	Unstandardized Regression Coefficients			
	U.S. Data Set		*Data for Other Countries*	
Independent Variable	*Main Effect, Interviews*	*Interaction, Interviews by Year*	*Main Effect, Interviews*	*Interaction, Interviews by Year*
Contacts (interview or questionnaire, logged)	.935	.954	.746	.712
Salience	.279	.269	.336	.344
Data collection by interview	.153	.182	.216	.342
Special third contact on questionnaire	.097	.095	.185	.191
Incentive on questionnaire	.145	.145	.154*	.154*
Government organization	.264	.259	.359	.360
General population	−.135	−.132	−.010*	−.013*
Post-1970 field work	—[a]	.038*	—[a]	−.001*
Interaction, interview by year	—[a]	−.358	—[a]	−.167*
Constant	1.158	1.160	.929	.925
R^2	.653	.658	.560	.567

[a] Inapplicable

*$p > .05$

Scoring notes: Response (proportion responding, after arcsine transformation—see footnote 3 for definition of response); number of interview calls or questionnaire contacts (frequency count—logged); salience (0 = not salient, 1 = possibly salient, 2 = salient); form of data collection (interviews = 1, questionnaires = 0); special third contact (0 = none, 1 = regular mail, 2 = special mail, 3 = telephone or personal contact); incentive on questionnaire (0 = none, 1 = less than 25¢, 2 = 25¢, 3 = 50¢, 4 = $1.00 or more); government organization, general population, post-1970 field work (dummy variables, scored 0,1).

about 30 percent. The intercept, of course, predicts response with all independent variables in the model equal to zero—here, that entails a questionnaire survey allowing only one contact (thus, no special follow-ups), a nonsalient topic, no provision for cash incentives, nongovernmental sponsorship, but with the offsetting advantage of a nongeneral population. The main effects model for the U.S. reveals that successive contacts would add .935 to the transformed response variable. For both mailed questionnaires and for interviews, total number of contacts is logged. The rationale is that successive calls are destined to yield diminishing returns, as convertible and initially not-at-home subjects are absorbed, and hard core refusals rise proportionately (Stinchcombe, et al., 1981: 383-4). Admittedly, the arcsine transformation on response presupposes that any response maximization technique, including interview callbacks and questionnaire follow-ups, yields diminishing returns at the extreme response ranges, but the intention in log transforming number of contacts is to model the reality that even in the middle range of survey response successive contacts do not yield constant increments in response.

Thus, the U.S. main effects model predicts that a second contact would raise *A* by .28 (log 2 × .935), equivalent to an increment of .13 in proportion responding. A third contact—that is, an initial mailing plus two follow-ups—would raise *p* to about .52. Were the topic on this hypothetical survey to be a "possibly salient" issue, projected response would rise by $A = .279$, yielding a predicted total response rate of 65 percent.

The coefficient of $b = .153$ showing the effect of interview data collection on *A*, tests hypothesis 1. Net of number of calls, salience of topic, and the other factors in the model, data collection by personal interview on average results in some 7.5 (again, derived from back-transformation) additional percentage points response over mailed questionnaires achieving returns in the 40–60 percent range. If a questionnaire achieved 80 percent response, the model predicts 6 additional percentage points response on the equivalent interview survey. As noted earlier, the transformation narrows the response margin near 100 percent; given 95 percent mailed response, for example, the prediction becomes 98 percent of attempted interviews successful. The coefficient for form of data collection is .05 significant, as a main effect, suggesting that hypothesis 1 should be upheld.

Though statistically significant, the advantage due to interviewers perhaps seems low,[6] in relation to the textbook estimates reviewed earlier. Apart from averaging a crossed interaction effect (discussed below), the small net effect reveals, we suspect, that interview surveys typically have involved more contacts than questionnaire surveys, probably because the prospects of success on interview callbacks have appeared brighter. A "not-at-home" on an interview enumeration is, unless deliberately evading the fieldworker, not necessarily reluctant to co-operate; assuming adequate postal delivery, non-respondents on the first mailing of a questionnaire are. The contribution of questionnaire experts such as Dillman (1974, 1978) has been to institutionalize for mail surveys the rigor of follow-up that was becoming standard procedure for probability sample interview surveys at least since the 1940s (Gaudet and Wilson, 1940). The 7.5 point net effect plausibly bridges the gap between the 70–75 percent questionnaire returns regarded as normal by Dillman and the higher rates "achievable by personal interview" (1974:755).

---

[6] Artifactually low, argued one reviewer, reasoning that the control for "general" population ignores the fact that the general populations typically surveyed by mail might be smaller and more rural than those covered in interviews. The argument can be addressed by noting: (1) Median *N* per observation on general populations is 947 for interviews, 800 for questionnaires; (2) 1.2 percent of interview surveys, against 3.9 percent of questionnaires, were collected exclusively in rural areas; such a tiny proportion, and difference in proportions, is not worth its degrees of freedom, as a control in the model; (3) expanding the model reported in Table 3 to include all categories of type-of-population (and sponsorship) raises the bonus for interview data collection to a maximum of about a percentage point of response, while encumbering the model with several nonsignificant coefficients.

The second model in Table 2, entering the dummy variable for post-1970 surveys as main effect and in interaction with form of data collection, tests hypothesis 2. The interaction is significant in the U.S. data, and reveals the hypothesized pattern of a larger advantage of interviews over questionnaires in the surveys conducted before the 1970s.[7] Indeed, as a net effect, the differential reverses in the post-1970 data. A survey providing, for example, for a total of five contacts, dealing with a "possibly salient" topic administered to a general population, would, from Table 2, score a predicted post-1970 return of 71 percent using a mailed questionnaire against 63 percent with personal interviews. The latter estimate approximates the figure reported by Koenig, et al. (1977:433) for recent interview surveys of heterogeneous populations (also, American Statistical Association, 1974:31). The model would, of course, acknowledge the likelihood of higher interview completion rates on maximum salience topics or after further callbacks.

A replication of the hypothesis tests appears in the righthand portion of Table 2. The net advantage to the interview method—hypothesis 1—reaches a (back-transformed) maximum of about 11 points in the non-U.S. data set of response rates, not significantly (at .05 level) different from the 7.5 point spread computed from the U.S. data. The agreement in other coefficients is also reasonably close, considering the alternative sampling frame employed for the Canadian portion of the non-U.S. data. The effect of the use of cash incentives does not reach significance in the non-U.S. data set but (Table 1) the case base for the replication is considerably smaller than for the U.S. data, and the nonsignificant coefficient is consistent as to direction with the first test. The trivial effect, in the replication, of general population is best interpreted cautiously. When computed for the Canadian replication alone, the negative effect of general population is equivalent to that in the U.S. data.

Hypothesis 2 is not formally upheld in the .05 level significance test for the non-U.S. data, though the direction of the interaction effect is consistent with the U.S. data. It may be justifiable to suspect Type II error in the second test, in light of the significant result already detected in the U.S. data. The interaction is much reduced in the non-U.S. data, however, and does not include the post-1970 reversal in advantage of interviews over questionnaires, observed in Table 2 (effect = .342 − .001 − .167 = +.174). The decline in interview response is not unique to the U.S., we conclude (see also, Market Research Society, 1976), but is particularly severe there.

---

[7] The nonsignificance of the main effect for year is unimportant, since the main and interaction effects are highly redundant (Cohen and Cohen, 1975:295). Calculated as a set, year and year-by-form-of-data-collection pass significance.

The data for the U.S. and for foreign surveys were next merged, and both unstandardized and standardized regression coefficients computed for the interaction model, with results entered in Table 3. The intention was to achieve a general statement about the relative importance of predictors of response. The unstandardized coefficients employed in earlier computations are, given inconsistent metrics, difficult to interpret when comparing size of effect across predictors within a model, particularly for the logged variable. In the standardized coefficients is re-expressed the notion that frequency of contact (b = .4) outweighs the importance of form of data collection (b = .16). As a net effect, use of personal interviews is not the dominant determinant of response implied in uncontrolled comparisons of interviews and questionnaires, so much as one of a group of factors, all of which are at least as important as form of collection. This reality is seen particularly strikingly in the cumulative R^2 tabulation in Table 3. Form of data collection year, and year-form interaction were deliberately entered as the first set in a hierarchial model, and explained some 4 percent of variance in response. The addition of the six remaining predictors raised R^2 to 62.5 percent. The residual 37.5 percent of variance is, as in any such computation, to be attributed to measurement error and to unmeasured variables; more elaborate interaction arrangements of the variables already in the model are, as noted earlier, unproductive. The interaction term in Table 3 falls shy of .05 significance but was retained for its reproducibility across two

Table 3. Summary Model Estimating the Importance of Form of Data Collection, Total Data Set (Dependent Variable: Arcsine Transformation of Proportion Responding)

Independent Variable	*Regression Coefficients*		*Cumulative R^2*
	Unstandardized	*Standardized*	
Data collection by interview	.206	.157	
Post-1970 field work	−.059*	−.045*	
Interaction, interview by year	−.143*	−.068*	.037
Number of contacts (logged)	.856	.436	
Salience	.272	.338	
Special third contact on questionnaire	.130	.193	
Incentive on questionnaire	.148	.142	
Government organization	.260	.188	
General population	−.099	−.084	.625
Constant	1.146		

* $p > .05$.

Scoring: As in Table 2.

data sets and because the intention in the present computation was to compare the maximum-estimate importance of form of data collection with the effect of other determinants of response.

Admittedly, the standardized coefficients and R^2 tabulations pose ambiguities for dummy variables, in the sense that the relative strength of predictors is a function of both size of effect (as measured in metric units) and the variances of independent variables (Cohen and Cohen, 1975: 177-78). After additional computations, weighting the form of data collection variable to fifty-fifty proportions, and creating sheaf coefficients for the sponsorship and type of population variables, we concluded that the standardized coefficients, when read in conjunction with the unstandardized, serve to underline our argument about the net effect of data collection by face-to-face interview.

Conclusion

For the first generation of survey researchers, the relative merits of data collection by mail or by face-to-face interview were straightforward. Questionnaires were acknowledged to cost little, to ensure standardization of question wordings, and possibly to encourage candor on sensitive probes (Reuss, 1943). The dismal response anticipated from all but the most motivated people was a decisive liability, however, and threatened a sociology-of-the-cooperative subject, descriptive of only a minority of most survey populations. If the price in fieldworkers' salaries could be borne, the interview promised high response as well as the opportunity to field long and complicated schedules.

Today, two trends have altered this "conventional wisdom" (Alwin, 1977:137): the perfection of response maximization techniques for mailed questionnaires, and the temporal decline in interview response rates. As the literature on maximizing questionnaire returns developed, it became increasingly evident that questionnaire response into the 70 percent range was attainable, even with general populations, and without a highly salient research topic. The great difference portrayed in many textbook reviews as separating interview and questionnaire response rates for general populations can from the results presented herein now be seen as partially the consequence of more zealous follow-ups on interview surveys. (Shosteck and Fairweather, 1979, have reached a similar conclusion, examining the case of research on physicians.) Thus, our position is that response on questionnaires and on interviews is essentially reducible to a common model, but that "inputs" into that model have traditionally been greater on interviews. Each form of data collection is highly elastic according to follow-ups and callbacks, at least until the 90–100 percent response range.[8] General populations are as detrimental to response on questionnaires as on interviews. In each mode of collection, government sponsorship is a legitimating force giving a marked increment in response. And the effect of salience of topic is indistinguishable for interviews and for questionnaires. Admittedly, questionnaires permit the employment of techniques such as cash incentives and special forms of contact, and that flexibility means that slightly more variance in questionnaire response can be accounted for than variance in interview response. Interviewers have traditionally eschewed such tactics, but by choice only. In the climate of diminished completions, interview survey researchers may, like psychologists recruiting subjects for experiments, begin to pay for information. The interview has at times (e.g., Champion and Sear, 1969) been employed as a form of final follow-up in questionnaire studies, and conceivably (see Dillman, 1978:289) it would be fruitful to reverse the procedure, contacting recurrent not-at-homes by mail, using a self-enumerated schedule.

The claim is not that there is nothing to be learned from comparing models of response for interviews and for questionnaires. As noted below, a finer-grained data set, allowing a distinction between failure to contact versus refusal, no doubt would be fruitful. What our computations are saying is that in data where components of total nonresponse are merged, the models of response for interviews and questionnaires turn out to be rather similar, even though the refusal component is probably the largest factor in questionnaire nonresponse, with interviews characterized by relatively large nonresponse through failure to contact.

To underline the second main conclusion of this paper, the results express quantitatively the profound importance for survey data collection of the social changes between the 1950s and 1980s. In the U.S. data set, the decline in interview response—callbacks, saliency, etc., controlled—amounts to $A = -.320$, equivalent to some 13 points, if the pre-1970 norm for interviews is set at about 85 percent. In the sample composed of Canadian and European observations, the decline is $A = -.168$—giving a loss of about 6 percentage points on an 85 percent response baseline. Despite concern about saturation surveying, questionnaire response has so far withstood the erosion, and even gained marginally in the U.S., net of enhanced follow-up practice. One key post-World War II demographic trend, the ascending educational floor, has worked *in favor* of the mailed-questionnaire researcher.

---

[8] Reviews of telephone survey response are beginning to reveal considerable heterogeneity in response rate for that method too, along the lines familiar to interview and questionnaire surveyors (Dillman, et al., 1976:67–8; Cummings, 1979:239; Groves and Kahn, 1979).

The reasons for declining interview response have received much attention elsewhere. A code for year of survey indexes what is likely a set of changes: saturation of the surveyed public; social disintegration arising from events in the 1960s; civil libertarian concerns about computerized information storage-linkage; accessibility problems due to a shift from detached housing to controlled-access multiple dwelling units and from high female labor force participation (American Statistical Association, 1974:31–32; Daniel, 1975:293–94; Market Research Society, 1976), with many of these trends proving particularly severe in urban areas (Steeh, 1981:46–50; Fitzgerald and Fuller, 1982:4).[9] The present design contributes to this literature by contrasting the consequences of social change for interview versus questionnaire response. The limitation to this contribution, however, is that as presently constituted, the interaction of form of data collection by time combines the two basic components—change over time in interviewers' ability to achieve contact, and a possible change in the persuasiveness of personal contacts. Our results do not confirm that, *given contact*, a personal appeal for survey cooperation is no longer as effective as a postal appeal. Probably the interviewer still holds a real though eroded advantage.

In sum, the difference in response due to form of data collection is fluid, and has never been as large as some uncontrolled comparisons have implied. The interview was superbly suited to an affluent industrial society settling into the easily enumerated suburban developments of the early postwar years. The traditional sexual division of labor was still sufficiently pronounced to guarantee a corps of readily accessible housewife respondents and appointment-makers. Educational levels remained low enough that oral enumeration posed a decided benefit, and labor costs for multiple calls by fieldworkers were modest. Interviews will always be indispensable for certain kinds of population, topic, and question type, and are proving valuable in combination strategies (Mangione, et al., 1982). The mailed questionnaire, however, and electronic variants such as the telephone survey and direct video interaction, are perhaps the optimal methods for surveying postindustrial society; these methods are tailormade for reaching a socially disintegrated citizenry, ensconsed in high-rise urban fortresses in which face-to-face contact has been delegitimated, yet still susceptible to the behavioral psychology of a carefully orchestrated series of impersonal contacts and follow-ups by survey researchers.

---

[9] We attempted to devise a city size variable, but only the crudest of codings could be achieved. Surveys conducted exclusively in urban areas were distinguishable from rural surveys and from urban-rural mixtures—a variable too crude for testing what might have been interesting three-way interactions such as number of contacts by form of collection by city size.

Appendix 1

Percentage Distributions on Variables Used in Nonresponse Analysis: Combined Data Set

	Interviews	*Questionnaires*
1. Proportion responding		
0 – .1999	1.0	5.4
.20 – .3999	11.6	20.7
.40 – .5999	19.6	25.4
.60 – .7999	35.7	23.6
.80 – 1.00	32.1	24.9
2. Total number of contacts		
one	18.8	40.7
two	14.3	25.4
three	16.1	18.9
four	11.6	10.1
five	8.0	2.6
six	9.8	0.8
seven	4.5	0.5
eight or more	9.0	0.3
missing	8.0	0.8
3. Salience of topic		
Not salient	18.8	20.7
Possibly salient	63.4	47.9
Salient	17.9	28.2
Missing	0.0	3.1
4. Form of data collection—see Table 1		
5. Special third contact on questionnaire		
None	—[a]	69.2
Regular mail	—[a]	18.9
Special mail	—[a]	4.4
Telephone or interview follow-up	—[a]	7.5
6. Incentive on questionnaire		
None	—[a]	87.8
Less than 25¢	—[a]	1.6
25¢	—[a]	2.6
50¢	—[a]	0.0
$1.00	—[a]	4.0
Other	—[a]	7.0
7. Sponsor		
Market research company	3.6	17.6
Government agency	16.1	20.2
University researchers	75.9	54.1
Institution (e.g., a charity)	4.5	6.7
Missing	0.0	1.3
8. Population		
General	72.3	19.9
Employee	0.0	3.4
School-army	0.9	6.5
Other organization	2.7	45.9
Other nongeneral population	24.1	23.6
Missing	0.0	0.8
9. Year of study—recoded into decades		
1920 – 29	0.0	1.0
1930 – 39	0.9	6.0
1940 – 49	12.5	12.7
1950 – 59	26.8	29.8
1960 – 69	27.7	30.6
1970 – 79	28.6	19.7
1980s	3.6	0.3

[a] Inapplicable.

For coding specifications, see Heberlein and Baumgartner (1978), Goyder (1982b), Missing data in regression models are handled by plugging means.

References

Alwin, D. F.
1977 "Making errors in surveys: an overview." Sociological Methods and Research 6:131–50.
American Statistical Association
1974 "Report on the ASA conference on surveys of human population." American Statistician 28:30–34.
Babbie, E. R.
1973 Survey Research Methods. Belmont CA: Wadsworth.
Bailey, K. D.
1978 Methods of Social Research. New York: Free Press.
Black, J. A., and D. J. Champion
1976 Methods and Issues in Social Research. New York: Wiley.
Bradburn, N. M., and S. Sudman
1979 Improving Interview Method and Questionnaire Design. San Francisco: Jossey-Bass.
Cannell, C. F., and R. L. Kahn
1968 "Interviewing." Pp. 526–95 in G. Lindzey and E. Aronson (eds.), The Handbook of Social Psychology, Vol. II (2nd ed.). Reading, MA: Addison-Wesley.
Champion, D. J., and A. M. Sear
1969 "Questionnaire response rates: a methodological analysis." Social Forces 47: 335–39.
Cohen, J., and P. Cohen
1975 Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences. Hillsdale, N.J.: Lawrence Erlbaum.
Cole, S.
1980 The Sociological Method: An Introduction to the Science of Sociology (3rd ed.). Chicago: Rand McNally.
Cummings, K. M.
1979 "Random digit dialing: a sampling technique for telephone surveys." Public Opinion Quarterly 43:233–44.
Daniel, W. W.
1975 "Nonresponse in sociological surveys: a review of some methods for handling the problem." Sociological Methods and Research 3:291–307.
DeMaio, T. J.
1980 "Refusals: who, where, and why?" Public Opinion Quarterly 44:223–33.
Dillman, D. A.
1978 Mail and Telephone Surveys: The Total Design Method. New York: Wiley.
Dillman, D. A., J. A. Christenson, E. H. Carpenter, and R. M. Brooks
1974 "Increasing mail questionnaire response: a four-state comparison." American Sociological Review 39:744–56.
Dillman, D. A., J. G. Gallegos, and J. H. Frey
1976 "Reducing refusal rates for telephone interviews." Public Opinion Quarterly 40:66–78.
Eichner, K., and W. Habermehl
1981 "Predicting response rates to mailed questionnaires" (comment on Heberlein and Baumgartner, ASR, August 1978). American Sociological Review 46:361–63.
Fitzgerald, J. D., and S. M. Cox
1975 Unravelling Social Science: A Primer on Perspectives, Methods, and Statistics. Chicago: Rand McNally.
Fitzgerald, R., and L. Fuller
1982 "I hear you knocking but you can't come in: the effects of reluctant respondents and refusers on sample survey estimates." Sociological Methods and Research 11:3–32.
Gallegos, J. C.
1974 "An experiment in maximizing response to telephone interviews through the use of a preliminary letter, based on the principles of exchange theory." M.A. thesis, Washington State University.

Gaudet, H., and E. C. Wilson
1940 "Who escapes the personal investigator?" Journal of Applied Psychology 24: 773–77.

Goode, W. J., and P. K. Hatt
1952 Methods in Social Research. New York: McGraw-Hill.

Goyder, J.
1982a "Nonresponse: the opinion surveyor's dilemma." Queen's Quarterly 89:569–82.
1982b "Further evidence on factors affecting response rates to mailed questionnaires." American Sociological Review 47:550–53.
In press "Nonresponse on surveys: a Canada-U.S. comparison." Canadian Journal of Sociology, in press.

Groves, R. M.
1979 "Actors and questions in telephone and personal interview surveys." Public Opinion Quarterly 43:190–205.

Groves, R. M., and R. L. Kahn
1979 Surveys by Telephone: A National Comparison with Personal Interviews. New York: Academic Press.

Heberlein, T. A., and R. Baumgertner
1978 "Factors affecting response rates to mailed questionnaires: a quantitative analysis of the published literature." American Sociological Review 43:447–62.
1981 "Is a questionnaire necessary in a second mailing." Public Opinion Quarterly 45:102–8.

Jahoda, M., M. Deutsch, and S. W. Cook
1951 Research Methods in Social Relations. New York: Dryden Press.

Kanuk, L., and C. Berenson
1975 "Mail surveys and response rates: a literature review." Journal of Marketing Research 12:440–53.

Kephart, W. M., and M. Bressler
1958 "Increasing the responses to mail questionnaires: a research study." Public Opinion Quarterly 22:123–32.

Kidder, L. H.
1981 Selltiz, Wrightsman and Cook's Research Methods in Social Relations (4th ed.). New York: Holt, Rinehart, and Winston.

Kish, L.
1965 Survey Sampling. New York: Wiley.

Koenig, D. J., G. R. Martin, and L. H. Seiler
1977 "Response rates and quality of data: a re-examination of the mail questionnaire." Canadian Review of Sociology and Anthropology 14:432–38.

Kviz, F. J.
1977 "Toward a standard definition of response rate." Public Opinion Quarterly 41:265–67.

Labovitz, S., and R. Hagedorn
1971 Introduction to Social Research. New York: McGraw-Hill.

Levine, S., and G. Gordon
1958 "Maximizing returns on mail questionnaires." Public Opinion Quarterly 22: 568–75.

Linsky, A. S.
1975 "Stimulating responses to mailed questionnaires: a review." Public Opinion Quarterly 39:82–101.

Mangione, T. W., R. Hingson, and J. Barrett
1982 "Collecting sensitive data: a comparison of three survey strategies." Sociological Methods and Research 10:337–46.

Market Research Society
1976 "Report of a working party of the MRS research and development committee: response rates in sample surveys." Journal of the Market Research Society 18:113–42.

Meyers, L. S., and N. E. Grossen
1974 Behavioral Research: Theory, Procedure, and Design. San Francisco: W. H. Freeman.

Miller, D. C.
1977 Handbook of Research Design and Social Measurement. New York: David McKay.
Nachmias, D., and C. Nachmias
1976 Research Methods in the Social Sciences. New York: St Martin's.
O'Neil, M. J.
1979 "Estimating the nonresponse bias due to refusals in telephone surveys." Public Opinion Quarterly 43:218–32.
Orenstein, A., and W. R. F. Phillips
1978 Understanding Social Research: An Introduction. Boston: Allyn and Bacon.
Parten, M.
1950 Surveys, Polls, and Samples: Practical Procedures. New York: Harper and Brothers.
Reuss, C. F.
1943 "Differences between persons responding and not responding to a mail questionnaire." American Sociological Review 8:433–38.
Rogers, T. F.
1976 "Interviews by telephone and in person: quality of responses and field performance." Public Opinion Quarterly 40:51–65.
Ross, H. Lawrence
1963 "The inaccessible respondent: a note on privacy in city and country." Public Opinion Quarterly 27:269–75.
Sanders, W. B., and T. K. Pinhey
1983 The Conduct of Social Research. New York: Holt, Rinehart and Winston.
Scott, C.
1961 "Research on mail surveys." Journal of the Royal Statistical Society 124:143–205.
Shosteck, H., and W. R. Fairweather
1979 "Physician response rates to mail and personal interview surveys." Public Opinion Quarterly 43:206–17.
Singer, E.
1978 "Informed consent: consequences for response rate and response quality in social surveys." American Sociological Review 43:144–61.
Sjoberg, G.
1955 "A questionnaire on questionnaires." Public Opinion Quarterly 18:423–27.
Sletto, R. F.
1940 "Pretesting of questionnaires." American Sociological Review 5:193–200.
Steeh, C.
1981 "Trends in response rates, 1952–1979." Public Opinion Quarterly 45:40–57.
Stinchcombe, A. L., C. Jones, and P. Sheatsley
1981 "Nonresponse bias for attitude questions." Public Opinion Quarterly 45:359–75.
Sudman, S., and N. M. Bradburn
1974 Response Effects in Surveys: A Review and Synthesis. Chicago: Aldine.
Survey Research Center
1976 Interviewer's Manual (rev. ed.). Ann Arbor: The University of Michigan.
True, J. A.
1983 Finding Out: Conducting and Evaluating Social Research. Belmont, CA: Wadsworth.
Vigderhous, G.
1977 "Analysis of patterns of response to mailed questionnaires." Sociological Methods and Research 6:205–13.
Weisberg, H. F. and B. D. Bowen
1977 An Introduction to Survey Research and Data Analysis. San Francisco: W. H. Freeman.
Young, P. V.
1949 Scientific Social Surveys and Research: An Introduction to the Background, Content, Methods, and Analysis of Social Studies. New York: Prentice-Hall.

The Polls:
America's Most Important Problems
Part I: National and International

TOM W. SMITH

In Memory of George Gallup, Sr.

ONE OF George Gallup's greatest achievements, as Burns Roper noted recently, was building time series by replicating questions. Among the hundreds of trends that Gallup collected, the most important problem series is unique. Gallup started asking about the most important problem facing the American people in September 1935. The trend line is now approaching the half-century mark and is as old as national polling itself. The most important problem item has appeared in at least 178 Gallup surveys since 1935, making it the second most frequently asked attitude after presidential popularity. This density allows not only the study of long-term trends in public opinion, but also the assessment of short-term reactions to specific events. Perhaps no other single item gives us as deep an understanding of American history over the last five decades. The most important problem question provides a grand overview of social change, describes history from the perspective of the participants, and helps to define distinct historical periods and identify turning points.

The standard Gallup usage asks, "What do you think is the most important problem facing this country today?" (See below, Question Wording: Most Important Problem, for the variations.)[1] The frame of reference is the country as a whole and elicits mentions of national and international problems rather than of local or personal problems. The question calls for a relative ranking of problems, not an absolute measure of the level of concern about problems in general or a particular problem. The question does not indicate whether the absolute level of concern has changed over time, but measures how the priority of problems has altered. In addition to the national most important problem question Gallup has

---

[1] Most impottant problem questions have also been asked by such survey organizations as NORC, Roper, SRC-Michigan, and Harris, but these are not included in Part I.

America's Most Important Problems, Part II: Regional, Community, and Personal will appear in the Fall 1985 issue of POQ.

Tom W. Smith is Senior Study Director, National Opinion Research Center (NORC), University of Chicago.

Public Opinion Quarterly Vol. 49:264–274 © by the Trustees of Columbia University
Published by Elsevier Science Publishing Co., Inc. 0033-362X/85/0049-264/$2.50

asked many close and distant variations. We excluded variants dealing with topics such as top campaign issues, problems the Congress or president should take up, questions one would ask the president or a member of Congress, what one would write Congress about, reasons for voting for or citicism of Republicans and Democrats, ideas for party platforms, desired news headlines, greatest war problem, and problem anticipated in the next four years.

On most surveys multiple responses were permitted. While the amount of multiple mentions varies, it typically falls in the 15–25 percent range. When multiple responses were allowed, we used number of mentions as the base and percentaged the figures to sum to 100 percent. We have used only 11 large categories to classify responses since changes in historical frames of references and coding schemes make finer codes unreliable. The four main groups are *Foreign affairs*, *Domestic*, *Not classified*, and *Don't knows*. *Foreign affairs* includes military and defense matters, demobilization, and space and is subdivided into Vietnam and other foreign affairs. *Domestic* covers economics, social control, civil rights, government, energy, and miscellaneous. *Economics* includes mentions of inflation, unemployment, job concerns, farms, trade, and labor. *Social control* includes violence, riots, crime, juvenile delinquency, drugs, moral decay, and the lack of religion. *Civil rights* covers race (excluding race riots) and some civil liberties issues. *Government* encompasses corruption, inefficiency, red tape, big government, and political/partisan criticism of leaders. *Energy* covers complaints about oil supplies, OPEC, embargoes, and hoarding, but not mentions of the high price of energy. *Miscellaneous* concerns named most frequently include poverty, slums, health, the environment, education, and transportation. *Unspecified* are unnamed miscellaneous responses. *Unclassified* items consist of topics that could not be placed in either the foreign or domestic sectors (most frequently the response "communism").

General Trends

The most important problem series reveals major shifts in public concerns over the last five decades. During the Great Depression the economy was the top worry. Early in 1939, however, foreign problems began to gain ground, and after Hitler's invasion of Poland foreign problems dominated public attention. With the close of World War II, problems about reconversion to a peacetime economy, labor unrest, and fears of a return of the Depression made economics the most mentioned problem. As the Cold War developed in the late 1940s, more and more attention shifted back to foreign affairs. From that point until the early 1960s, foreign affairs dominated public concern, ranking first in 48 of 56 surveys and often commanding over 50 percent of the public. Economics remained a strong second during the 1950s and overcame foreign affairs six times during the recessions of 1949, 1954, and 1958. In 1963 the hegemony of foreign affairs was interrupted by the emergence of the civil rights movement. Civil rights ranked as the most important problem for most of the next two years until foreign affairs, boosted by the Vietnam War, regained the top position in 1965. From 1965 to 1970 Vietnam and other international issues dominated public concern. The only exception occurred in August 1967, when race riots pushed social control to the forefront.

Over the following two and a half years (1970–1973), concerns were very unstable. Foreign affairs topped the list seven times, followed by social control with five, and the economy with four. In addition, mentions of topics in the miscellaneous category such as welfare concerns (poverty, slums, education, health care, etc.) and environmental issues were at unusually high levels during this period. Finally, in the spring of 1973 public concern began to coalesce around economics. Since then economics has ranked first in all but two instances—January 1973, when the Arab oil embargo shot energy to the top, and January 1980, when the Iranian hostage crisis and the Soviet invasion of Afghanistan pushed foreign affairs to the head of the list. With these minor exceptions, economics has completely dominated public concerns for the last 10 years, often capturing over 60 percent of the public. The level of concern rarely dropped below an absolute majority and typically fell below the 50 percent mark only when energy concerns periodically peaked at 10 percent or higher.

As interesting as the switches in most important problem are the separate trends for the major topics—foreign affairs and Vietnam, economics, civil rights, social control, the government, and energy. Foreign affairs was often the top problem and we have already described much of its trend. International concerns rose with the start of the Cold War and except for breaks during recessions dominated the problem list until 1963. Concern then dropped off in 1963–65 as the civil rights movement gained attention. With the escalation of the Vietnam War, however, foreign affairs regained the top position in 1965. Worries about the Vietnam War fluctuated around a high level from late 1966 to early 1968 and then began to decline until they virtually disappeared as an issue by mid-1973. As Vietnam disappeared as an issue so did foreign affairs. Like Mr. Podsnap, the public seems to have swept international problems behind it with a wave of its arm. With the exception of the upsurge over the hostage/Afghanistan concern in January 1980, the public kept foreign affairs near the bottom of the list until late 1982. Since then, foreign affairs, centering around the issue of nuclear war and superpower rivalries, has begun to recapture public attention.

Economic concerns are closely tied to objective economic conditions and sensitive to fluctuations in the business cycle. Concerns were high during the Great Depression, dropped off sharply as World War II started in Europe and sprang back to first place at the close of World War II. From then to the mid 1960s economic worries showed a long-term decline, while on the short term they followed the business cycle with peaks during the recessions of 1949, 1954, 1958 and 1960–61. During the mid to late 1960s economic concerns remained at historic lows. But in 1970 declining growth and increasing inflation began to boost economic worries. By 1973 a majority started to mention economic problems and, with the few exceptions noted above, economics has continued to dominate public attention to the present.

Before 1956 almost no one considered civil rights as the most important problem. As Ronald Reagan noted during the 1980 presidential debate, it was a time "when this country didn't even know it had a racial problem." The Montgomery bus boycott, Little Rock school desegregation confrontation, and later struggles over desegregating restaurants and other public facilities sharply, but briefly, stirred up concern from 1956 to the spring of 1960. Concern then subsided until early 1963 (although remaining well above the minuscule levels registered before 1956). Increased agitation by the civil rights movement, hallmarked by Martin Luther King's "I have a dream" speech at the Washington Monument in August, triggered a surge in public concern that was both larger and more sustained than earlier peaks. From 1963 until 1965 civil rights regularly topped the problems list. Preempted by Vietnam and social disorders, civil rights began to decline in 1965. Since 1974 civil rights has virtually vanished as a public concern and has fallen back to pre-1956 levels.

Social control problems rarely made a ripple until the late 1940s. Then fears about Communist subversion—generated by the Cold War, celebrated spying cases in the United States and England, and Red hunting investigations led by Senator Joseph McCarthy—peaked concern in the spring of 1954. When worries about domestic Communists then subsided, social control concerns were reduced to minor mentions of juvenile delinquency and other criminal matters. Low concern about social control continued until 1965–66, when race riots, a rising crime rate, and increasingly violent protests began to drive up worries. Peaks were reached in 1967 following major race riots, in the fall of 1970 after a summer of race and campus riots and an escalation in political violence, and in October 1971, when a variant wording and coding scheme increased references to this problem. With the end of campus protests and major race riots, concern about social control fell off to less than half its peak level by 1974. Since 1974 social control has remained at a low level consisting mostly of fear of crime.

Government has never been ranked as the top problem, and before 1973 was rarely named as a problem. Watergate greatly increased mentions of this issue and it reached its high point of 23 percent just prior to Nixon's resignation. With Nixon's departure government declined once again into the background as a minor issue.

Energy emerged as a national concern with the Arab oil boycott of 1973–74. It averaged 2 percent in early 1973 before surging to 34 percent in January 1974. With the end of the boycott, energy quickly fell back to less than 5 percent. Energy showed high points of concern during the winter heating seasons in 1977, 1978, 1979, and throughout 1979 as the second major round of oil price hikes went into effect. After early 1980 concern subsided and practically disappeared after late 1981.

The American Institute of Public Opinion (Gallup Poll) collected the data, which were obtained from the Roper Center, University of Connecticut.

Question Wording: Most Important Problem[2]	*Surveys*
A. What do you regard as the most vital issue before the American people today?	AIPO (9/19/35), 8, (12/36), 62, 106
B. What do you regard as the most important problem before the American people today?	AIPO143, 155
C. What do you think is the most important problem before the American people today?	AIPO176, 205, 447
D. What do you think is the most important issue before the country today?	AIPO252
E. In your opinion, what is the most important problem the United States government must solve in the next few months?	AIPO254
F. Aside from winning the war, what do you think is the most important problem facing this country today?	AIPO286, 342
G. Which problem here in the United States have you, yourself, been thinking about most during the last few days?	AIPO298

---

[2] Several of these items (e.g., I, K, Q, R, T) are known to produce different distributions because of question wording variation. They have been included in order to document all versions of the question and to demonstrate the variation that occurs.

	Question Wording: Most Important Problem	*Surveys*
H.	What do you think will be the most important problem facing this country during the next year?	AIPO354, 357, 374T
I.	In your opinion, what is the most important problem the government must solve in the next few months?	AIPO366
J.	What do you think will be the most important problem facing this country in the next six months?	AIPO374K
K.	In your opinion, what is the most important problem the United States (U.S.) government must solve in the next year?	AIPO387, 449
L.	What do you think is the most important problem facing this (the) country today?	AIPO 401, 404, 415, 420, 435, 443T, 494, 528, 570–664 669–836, 840–1131
M.	What do you think is the most important problem facing the entire country today?	AIPO454, 480
N.	What do you regard as the biggest issue or problem facing the U.S. government today?	AIPO530
O.	What do you regard as the biggest issue or problem facing the government in Washington today?	AIPO548, 555
P.	Of the many problems facing this country, which one gives you the greatest worry or concern at this time?	AIPO665
Q.	Of the many problems which this country has, which one gives you the greatest worry or concern at this time?	AIPO666
R.	What are your chief worries these days—what things bother or upset you most? Now, what do you fear most?	POS655
S.	Now, how about the United States. Of ALL the problems—domestic and international facing this country today, which ONE do you think is the most important?	POS655
T.	Turning to the next topic . . . what bothers or angers you the most about America today?	AIPO839

Most Important Problem Marginals (Proportion)

Survey	AIPO	AIPO	AIPO106	AIPO143	AIPO155	AIPO176	AIPO205	AIPO252
Date	(9/35)	(12/36)	(12/37)	(1/39)	(4/39)	(11/39)	(8/40)	(11/41)
Foreign affairs	.110	.240	.270	.134	.349	.468	.523	.806
Domestic	.890	.760	.660	.785	.607	.477	.422	.165
Economic	.610	.580	.460	.596	.494	.325	.265	.125
Social control	.040	.020	—	.022	.014	.018	.028	.010
Civil rights	—	—	—	.006	.002	.001	.001	.001
Government	.110	.020	.050	.074	.021	.040	.068	.009
Misc., named	.110	.140	.030	.087	.069	.093	.050	.016
Unspecified	.020	—	.120	—	.007	—	.009	.004
Not classified	—	—	—	—	.001	.004	—	—
Don't know	—	—	.070	.080	.044	.050	.057	.029
N cases	N.A.[a]	N.A.	N.A.	3063	3122	3192	3094	3065
N responses	N.A.	N.A.	N.A.	3063	3122	3192	3094	3232
Responses/cases	—	—	—	1.00	1.00	1.00	1.00	1.05

Survey	AIPO254	AIPO286	AIPO342	AIPO354	AIPO357	AIPO366	AIPO374	AIPO387
Date	(11/41)	(12/42)	(3/45)	(8/45)	(10/45)	(2/46)	(6/46)	(1/47)
Foreign affairs	.750	.283	.242	.113	.121	.273	.136	.252
Domestic	.134	.615	.661	.884	.875	.726	.860	.745
Economic	.067	.493	.510	.847	.826	.594	.786	.609
Social control	—	.022	.060	.006	.008	.001	.004	.001
Civil rights	—	.001	.049	.014	.013	.003	.004	.006
Government	—	.016	—	.002	.002	.006	.010	.015
Misc., named	—	.083	.027	.015	.026	.117	.051	.106
Unspecified	.067	—	.015	—	—	.005	.005	.008
Not classified	—	—	—	.001	—	.001	.002	.002
Don't know	.117	.102	.097	.004	.004	—	.001	.000
N cases	N.A.	2935	3055	3106	3096	3122	3119	1468
N responses	N.A.	3151	3318	3402	3561	3400	3362	1546
Responses/cases	1.20	1.07	1.09	1.10	1.15	1.09	1.08	1.05

Survey	AIPO401	AIPO404	AIPO415	AIPO420	AIPO435	AIPO443T	AIPO447	AIPO449
Date	(7/47)	(9/47)	(3/48)	(6/48)	(1/49)	(6/49)	(9/49)	(11/49)
Foreign affairs	.461	.323	.676	.[illegible]7	.458	.222	.284	.317
Domestic	.464	.632	.228	.429	.492	.603	.591	.568
Economic	.354	.515	.096	.268	.347	.399	.371	.473
Social control	.014	.032	—	.024	.024	.008	.031	.001
Civil rights	.008	.008	—	—	.010	.028	.023	.009
Government	—	.006	—	—	.013	.018	.043	.018
Misc., named	.075	.046	.017	.045	.062	.146	.120	.064
Unspecified	.013	.025	.115	.092	.036	.004	.003	.003
Not classified	.014	.020	.065	.040	—	.008	.067	.025
Don't know	.061	.026	.032	.033	.050	.167	.059	.092
N cases	3006	3056	2894	3154	3138	2765	2896	2903
N responses	3210	3269	2926	3197	3267	3311	3029	3087
Responses/cases	1.07	1.07	1.01	1.01	1.04	1.20	1.05	1.06

Survey	AIPO454	AIPO480	AIPO494	AIPO528	AIPO530	AIPO548	AIPO555
Date	(3/50)	(9/51)	(6/52)	(3/54)	(5/54)	(6/55)	(10/55)
Foreign affairs	.446	.562	.544	.345	.485	.483	.407
Domestic	.444	.404	.429	.603	.280	.247	.379
Economic	.289	.244	.263	.362	.186	.136	.204
Social control	.107	.039	.074	.178	.001	.005	.009
Civil rights	.010	.003	.004	.004	.009	.040	.025
Government	.030	.036	.062	.024	.028	.009	.054
Misc., named	.058	.043	.008	.006	.013	.016	.016
Unspecified	.034	.039	.017	.028	.043	.041	.071
Not classified	—	.008	—	—	.142	.058	.045
Don't know	.026	.027	.026	.051	.093	.213	.169
N cases	1458	1986	2031	1562	1415	1462	1577
N responses	1525	2093	2258	1752	1524	1524	1662
Responses/cases	1.05	1.05	1.11	1.12	1.08	1.04	1.05

Survey	AIPO570	AIPO571	AIPO572	AIPO573	AIPO583	AIPO588	AIPO590	AIPO593
Date	(9/56)	(9/56)	(10/56)	(10/56)	(5/57)	(8/57)	(10/57)	(12/57)
Foreign affairs	.390	.443	.451	.466	.431	.370	.383	.497
Domestic	.485	.426	.469	.434	.462	.506	.519	.405
Economic	.203	.189	.260	.253	.299	.232	.161	.210
Social control	.020	.030	.028	.020	.054	.050	.038	.047
Civil rights	.177	.121	.095	.081	.041	.176	.290	.045
Government	—	.004	.003	—	—	—	.001	.010
Misc., named	.035	.032	.052	.022	.046	.033	.014	.052
Unspecified	.050	.051	.013	.038	.023	.015	.015	.040
Not classified	.029	.032	.017	.020	.041	.025	.019	.001
Don't know	.093	.099	.080	.100	.066	.099	.079	.097
N cases	1979	2207	2223	2175	1570	1528	1558	1527
N responses	2332	2439	2465	2415	1570	1805	1558	1527
Responses/cases	1.18	1.11	1.11	1.11	1.00	1.19	1.00	1.00

(Continued)

Survey	AIPO596	AIPO601	AIPO603	AIPO604	AIPO605	AIPO606	AIPO610	AIPO612
Date	(3/58)	(7/58)	(8/58)	(9/58)	(9/58)	(10/58)	(2/59)	(3/59)
Foreign affairs	.253	.327	.480	.531	.527	.460	.439	.476
Domestic	.628	.524	.441	.405	.406	.461	.492	.461
Economic	.472	.371	.223	.164	.179	.240	.269	.277
Social control	.044	.023	.030	.018	.007	.013	.013	.025
Civil rights	.039	.059	.093	.165	.174	.139	.100	.067
Government	.004	.011	—	—	—	—	—	—
Misc., named	.024	.020	.010	.013	.006	.010	.039	.014
Unspecified	.045	.040	.084	.045	.040	.059	.071	.077
Not classified	.034	.052	—	—	—	—	—	—
Don't know	.085	.098	.078	.065	.067	.079	.068	.064
N cases	1609	1513	1563	1514	1665	1553	1616	1737
N responses	1609	1513	1563	1514	1665	1553	1616	1737
Responses/cases	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Survey	AIPO616	AIPO618	AIPO624	AIPO627	AIPO628	AIPO629	AIPO630	AIPO631
Date	(7/59)	(9/59)	(2/60)	(4/60)	(5/60)	(6/60)	(6/60)	(7/60)
Foreign affairs	.460	.547	.469	.401	.606	.527	.552	.569
Domestic	.477	.377	.420	.481	.320	.309	.291	.298
Economic	.316	.230	.183	.201	.143	.111	.127	.125
Social control	.013	.037	.028	.022	.012	.008	.027	.027
Civil rights	.078	.049	.052	.157	.052	.057	.054	.057
Government	—	—	.006	.007	.004	.001	.002	.002
Misc., named	.024	.013	.060	.030	.016	.009	.022	.017
Unspecified	.046	.049	.091	.063	.093	.122	.059	.071
Not classified	—	—	.063	.051	.035	.099	.093	.078
Don't know	.063	.076	.048	.067	.039	.065	.064	.055
N cases	1532	5778	3135	2759	3044	2519	3254	2789
N responses	1532	5778	3484	2900	3347	2779	3692	3275
Responses/cases	1.00	1.00	1.11	1.05	1.10	1.10	1.13	1.17

Survey	AIPO632	AIPO633	AIPO634	AIPO635	AIPO636	AIPO637	AIPO641	AIPO644
Date	(7/60)	(8/60)	(8/60)	(9/60)	(9/60)	(10/60)	([illegible]1)	(5/61)
Foreign affairs	.597	.578	.568	.587	.586	.551	.424	.467
Vietnam	—	—	—	—	—	—	—	.028[b]
Domestic	.275	.298	.279	.287	.269	.313	.433	.323
Economic	.140	.147	.132	.140	.158	.187	.335	.206
Social control	.020	.018	.026	.023	.017	.004	.017	.023
Civil rights	.062	.062	.068	.060	.035	.049	.051	.050
Government	.002	.005	.004	.003	.002	.003	.001	.001
Misc., named	.021	.024	.026	.035	.017	.018	.020	.034
Unspecified	.030	.041	.023	.025	.040	.052	.009	.009
Not classified	.089	.066	.088	.072	.090	.094	.094	.174
Don't know	.039	.057	.064	.054	.055	.041	.050	.036
N cases	3162	3077	3337	2906	3614	2988	2873	3545
N responses	3485	3542	3789	3273	4480	3461	3310	4415
Responses/cases	1.10	1.15	1.14	1.13	1.24	1.16	1.15	1.25

Survey	AIPO648	AIPO653	AIPO657	AIPO660	AIPO662	AIPO663	AIPO664	AIPO665
Date	(7/61)	(12/61)	(4/62)	(6/62)	(8/62)	(9/62)	(10/62)	(11/62)
Foreign affairs	.631	.598	.448	.349	.509	.649	.656	.722
Vietnam	—	—	.001	.006	.003	—	—	—
Domestic	.210	.241	.338	.437	.317	.299	.300	.166
Economic	.117	.112	.184	.234	.181	.130	.109	.081
Social control	.019	.018	.050	.058	.040	.007	.011	.01[illegible]
Civil rights	.038	.035	.055	.075	.048	.076	.108	.021
Government	.001	.002	.006	.003	.012	—	—	—
Misc., named	.011	.035	.020	.026	.021	.085	.072	.053
Unspecified	.023	.039	.022	.040	.015	—	—	—
Not classified	.119	.099	.130	.136	.110	—	—	.057
Don't know	.040	.062	.085	.078	.063	.053	.043	.056
N cases	3158	2988	3403	3275	3350	3938	4248	4426
N responses	3459	3409	3697	3759	3963	4163	4428	5265
Responses/cases	1.10	1.14	1.09	1.15	1.18	1.06	1.04	1.19

Survey	AIPO666	AIPO669[c]	AIPO677	AIPO688	AIPO689	AIPO694	AIPO695	AIPO696
Date	(12/62)	(3/63)	(9/63)	(3/64)	(4/64)	(6/64)	(7/64)	(8/64)
Foreign affairs	.609	.606	.245	.284	.279	.243	.202	.408
Vietnam	—	—	—	.018	.013	.060	.030	.103
Domestic	.268	.327	.686	.541	.600	.644	.682	.484
Economic	.124	.173	.133	.129	.126	.084	.064	.055
Social control	.033	—	.003	.013	.013	.009	.038	.038
Civil rights	.046	.038	.483	.304	.362	.419	.506	.304
Government	.012	—	—	—	—	—	—	—
Misc., named	.010	—	.067	.059	.060	.106	.050	.057
Unspecified	.043	.116	—	.035	.039	.025	.024	.029
Not classified	.056	—	—	.100	.077	.072	.064	.060
Don't know	.067	.067	.068	.075	.044	.041	.052	.047
N cases	3193	N.A.	3230	3539	3509	3506	3515	3513
N responses	3456	N.A.	3442	3943	4025	3922	4092	4214
Responses/cases	1.08	1.04	1.07	1.11	1.15	1.12	1.16	1.20

Survey	AIPO697	AIPO698	AIPO699	AIPO706	POS655	POS655	POS655	AIPO708
Date	(8/64)	(9/64)	(10/64)	(2/65)	(2/65) (Problems)	(2/65) (Worries)	(2/65) (Fears)	(3/65)
Foreign affairs	.280	.332	.408	.525	.517	.197	.258	.365
Vietnam	.054	.054	.052	.248	—	—	—	.198
Domestic	.612	.559	.524	.422	.426	.697	.478	.612
Economic	.088	.102	.120	.076	.047	.261	.094	.050
Social control	.039	.035	.062	.066	.011	—	.057	.067
Civil rights	.397	.290	.206	.217	.207	—	—	.449
Government	—	.034	.028	.009	—	—	—	.015
Misc., named	.026	.038	.091	.045	.020	.284	.182	.021
Unspecified	.061	.060	.015	.009	.141	.152	.145	.010
Not classified	.064	.057	—	—	—	—	—	—
Don't know	.044	.052	.069	.053	.057	.106	.263	.023
N cases	4003	3590	3503	3505	1757	1620	1620	3500
N responses	4683	4233	4157	4043	1757	2140	1726	4060
Responses/cases	1.17	1.18	1.19	1.15	1.00	1.32	1.07	1.16

Survey	AIPO711	AIPO714	AIPO715	AIPO717	AIPO719	AIPO720	AIPO728	AIPO733
Date	(3/65)	(7/65)	(8/65)	(9/65)	(10/65)	(11/65)	(5/66)	(8/66)
Foreign affairs	.532	.557	.572	.460	.550	.552	.569	.469
Vietnam	.191	.332	.349	.160	.312	.295	.423	.386
Domestic	.429	.413	.388	.499	.415	.391	.392	.508
Economic	.069	.071	.055	.079	.074	.071	.160	.153
Social control	.082	.080	.070	.081	.098	.074	.084	.120
Civil rights	.193	.181	.187	.226	.148	.171	.075	.167
Government	.007	.015	.017	.012	.024	.016	.029	.016
Misc., named	.042	.039	.040	.075	.059	.037	.038	.039
Unspecified	.036	.027	.018	.026	.011	.022	.006	.013
Not classified	—	—	—	—	—	—	—	—
Don't know	.039	.030	.040	.041	.035	.058	.042	.025
N cases	3546	1590	3527	3555	3525	3521	1563	1509
N responses	4220	1809	4030	4134	4116	3940	1706	1808
Responses/cases	1.19	1.14	1.14	1.16	1.17	1.12	1.09	1.20

Survey	AIPO735	AIPO736	AIPO740	AIPO749	AIPO753	AIPO756	AIPO761[c]	AIPO765[d]
Date	(10/66)	(10/66)	(1/67)	(8/67)	(10/67)	(1/68)	(1/68)	(7/68)
Foreign affairs	.544	.525	.582	.391	.524	.516	.385	.407
Vietnam	.450	.377	.459	.338	.432	.468	.385	.352
Domestic	.427	.434	.391	.581	.469	.472	.588	.584
Economic	.149	.178	.146	.071	.104	.122	.073	.087
Social control	.057	.056	.064	.408	.182	.159	.166	.294
Civil rights	.164	.135	.089	.055	.106	.108	.229	.111
Government	.014	.021	.024	.008	.029	.021	—	.038
Misc., named	.036	.032	.053	.029	.035	.048	.037	.044
Unspecified	.007	.012	.015	.010	.013	.014	.083	.009
Not classified	—	—	—	—	—	—	—	—

(Continued)

Survey	AIPO735	AIPO736	AIPO740	AIPO749	AIPO753	AIPO756	AIPO761[c]	AIPO765[d]
Don't know	.029	.041	.028	.029	.007	.012	.028	.010
N cases	1597	3510	3491	1627	1648	1502	N.A.	1526
N responses	1847	3979	4183	1627	1776	1681	N.A.	1938
Responses/cases	1.16	1.13	1.20	1.00	1.08	1.12	1.09	1.27

Survey	AIPO766	AIPO767	AIPO769	AIPO770	AIPO773	AIPO781	AIPO797	AIPO807
Date	(7/68)	(8/68)	(9/68)	(10/68)	(1/69)	(5/69)	(1/70)	(5/70)
Foreign affairs	.436	.378	.390	.408	.440	.427	.331	.318
Vietnam	.401	.349	.341	.383	.355	.384	.280	.193
Domestic	.552	.606	.592	.570	.543	.564	.655	.663
Economic	.079	.062	.076	.069	.095	.066	.157	.063
Social control	.225	.317	.305	.288	.177	.312	.221	.061
Civil rights	.171	.148	.147	.151	.195	.118	.112	.079
Government	.028	.032	.024	.019	.016	.004	.008	.019
Misc., named	.046	.037	.038	.034	.058	.052	.147	.190
Unspecified	.003	.010	.002	.010	.008	.013	.011	.251
Not classified	—	—	—	—	—	—	—	—
Don't know	.012	.016	.018	.022	.017	.009	.014	.019
N cases	1526	1507	1500	1605	1461	1539	1573	1509
N responses	1785	1827	1785	1845	1666	1569	1776	1694
Responses/cases	1.17	1.21	1.19	1.15	1.14	1.02	1.13	1.12

Survey	AIPO811	AIPO814	AIPO815	AIPO824	AIPO831	AIPO836	AIPO839	AIPO840
Date	(7/70)	(9/70)	(10/70)	(2/71)	(6/71)	(8/71)	(10/71)	(11/71)
Foreign affairs	.283	.327	.322	.338	.323	.220	.164	.197
Vietnam	.199	.214	.237	.254	.266	.181	.107	.129
Domestic	.696	.651	.656	.645	.664	.767	.786	.779
Economic	.140	.101	.110	.221	.116	.347	.127	.370
Social control	.322	.365	.407	.201	.292	.208	.404	.218
Civil rights	.082	.087	.054	.061	.058	.051	.061	.053
Government	.013	.015	.013	.011	.011	.022	.070	.023
Misc., named	.136	.074	.067	.140	.170	.119	.064	.108
Unspecified	.003	.009	.005	.011	.017	.019	.060	.008
Not classified	—	—	—	—	—	—	—	—
Don't know	.021	.021	.023	.016	.013	.015	.050	.024
N cases	1500	1497	1507	1571	1591	1547	1558	1558
N responses	1780	1814	1771	1731	1860	2037	1958	1851
Responses/cases	1.19	1.21	1.18	1.10	1.17	1.32	1.26	1.19

Survey	AIPO844	AIPO851	AIPO853	AIPO855	AIPO858	AIPO859	AIPO862	AIPO864
Date	(2/72)	(4/72)	(6/72)	(7/72)	(9/72)	(10/72)	(1/73)	(2/73)
Foreign affairs	.247	.301	.356	.251	.232	.308	.349	.091
Vietnam	.199	.233	.273	.207	.223	.247	.291	.032
Domestic	.737	.684	.620	.737	.696	.665	.629	.886
Economic	.273	.217	.214	.211	.233	.264	.254	.386
Social control	.252	.251	.192	.282	.253	.218	.215	.249
Civil rights	.042	.036	.044	.042	.037	.028	.038	.057
Government	.031	.037	.020	.038	.039	.034	.022	.024
Energy	—	—	—	—	—	—	—	—
Misc., named	.123	.126	.119	.128	.112	.106	.093	.164
Unspecified	.015	.017	.030	.037	.022	.015	.006	.006
Not classified	—	—	—	—	—	—	—	—
Don't know	.016	.016	.025	.012	.022	.027	.022	.022
N cases	1502	1542	1516	1526	1505	1516	1549	1517
N responses	1871	1918	1732	1751	1773	1848	1854	1740
Responses/cases	1.25	1.24	1.14	1.15	1.18	1.22	1.20	1.15

Survey	AIPO870	AIPO871	AIPO877	AIPO886	AIPO906	AIPO913	AIPO915	AIPO916
Date	(5/73)	(5/73)	(9/73)	(1/74)	(5/74)	(8/74)	(9/74)	(10/74)
Foreign affairs	.087	.060	.055	.039	.040	.045	.018	.017
Vietnam	.029	.034	.007	.004	.003	.001	—	—
Domestic	.887	.916	.918	.898	.925	.935	.965	.964
Economic	.397	.325	.569	.240	.457	.694	.748	.725
Social control	.215	.188	.117	.074	.123	.077	.067	.055
Civil rights	.032	.027	.013	.010	.009	.011	.006	.007
Government	.125	.199	.112	.176	.234	.081	.077	.099
Energy	.015	.020	.027	.342	.039	.008	.016	.020
Misc., named	.096	.088	.070	.031	.051	.045	.033	.040
Unspecified	.008	.069	.010	.025	.013	.019	.018	.019
Not classified	—	—	—	—	—	—	—	—
Don't know	.027	.023	.026	.064	.035	.020	.018	.019
N cases	1531	1548	1502	1589	1509	1590	1527	1586
N responses	1698	1790	1841	1996	1760	1834	1709	1827
Responses/cases	1.11	1.16	1.23	1.26	1.17	1.15	1.12	1.15

Survey	AIPO924	AIPO932	AIPO938	AIPO943	AIPO950	AIPO960	AIPO961	AIPO970
Date	(2/75)	(7/75)	(10/75)	(12/75)	(4/76)	(10/76)	(10/76)	(2/77)
Foreign affairs	.044	.031	.034	.067	.036	.079	.048	.090
Vietnam	.003	.001	—	—	—	—	—	—
Domestic	.956	.950	.942	.911	.942	.896	.923	.901
Economic	.612	.615	.648	.613	.509	.639	.650	.465
Social control	.137	.108	.109	.146	.161	.093	.106	.145
Civil rights	.005	.007	.017	.008	.029	.010	.011	.022
Government	.073	.096	.065	.054	.140	.087	.075	.034
Energy	.058	.048	.050	.022	.016	.010	.016	.104
Misc., named	.052	.039	.042	.043	.062	.049	.050	.099
Unspecified	.019	.038	.010	.024	.025	.009	.015	.032
Not classified	—	—	—	—	—	—	—	—
Don't know	—	.020	.024	.022	.022	.025	.029	.009
N cases	1576	1561	1553[a]	1572	1549	1538	1550	1525
N responses	2118	1871	1950	1929	2061	2074	2036	3300[b]
Responses/cases	1.34	1.20	1.26	1.23	1.33	1.35	1.31	2.16

Survey	AIPO980	AIPO986	AIPO993	AIPO999	AIPO1003[c]	AIPO1006	AIPO1011
Date	(7/77)	(10/77)	(2/78)	(4/78)	(5/78)	(7/78)	(9/78)
Foreign affairs	.081	.058	.068	.072	.083	.080	.045
Domestic	.861	.916	.903	.904	.894	.895	.931
Economic	.428	.498	.481	.621	.612	.643	.642
Social control	.118	.115	.095	.085	.058	.083	.105
Civil rights	.014	.021	.012	.012	.017	.017	.013
Government	.041	.024	.031	.034	.025	.043	.041
Energy	.136	.153	.203	.073	.033	.032	.043
Misc., named	.096	.090	.068	.060	—	.065	.073
Unspecified	.028	.015	.013	.019	.149	.012	.014
Not classified	—	—	—	—	—	—	—
Don't know	.058	.027	.029	.022	.025	.026	.025
N cases	1516	1517	1515	1546	1508	1555	1530
N responses	1811	1826	1785	1828	N.A.	1897	1747
Responses/cases	1.19	1.20	1.18	1.18	1.21	1.22	1.14

(Continued)

Survey	AIPO1023	AIPO1028	AIPO1036	AIPO1041	AIPO1047	AIPO1051	AIPO1057	AIPO1059
Date	(2/79)	(5/79)	(8/79)	(10/79)	(1/80)	(3/80)	(6/80)	(7/80)
Foreign affairs	.140	.041	.045	.056	.412	.144	.167	.104
Domestic	.844	.942	.937	.927	.575	.847	.809	.881
Economic	.576	.507	.465	.591	.368	.671	.621	.630
Social control	.064	.066	.151	.062	.050	.042	.047	.085
Civil rights	.004	.005	.008	.003	.004	.003	.004	.003
Government	.028	.047	.086	.039	.030	.035	.044	.067
Energy	.117	.273	.155	.178	.106	.072	.060	.055
Misc., named	.044	.036	.053	.039	.016	.020	.024	.036
Unspecified	.011	.008	.019	.015	.001	.004	.009	.005
Not classified	—	—	—	—	—	—	—	—
Don't know	.017	.017	.019	.018	.013	.009	.023	.014
N cases	1534	1511	1589	1539	1597	1571	1583	1548
N responses	1905	1897	2038	1852	1937	1870	1896	1863
Responses/cases	1.24	1.26	1.28	1.20	1.21	1.19	1.20	1.20

Survey	AIPO1062	AIPO1063	AIPO1064	AIPO1068	AIPO1073	AIPO1083	AIPO1088	AIPO1092
Date	(9/80)	(10/80)	(11/80)[e]	(1-2/81)[e]	(5/81)	(10/81)	(1/82)	(4/82)
Foreign affairs	.136	.134	.040	.056	.095	.108	.103	.104
Domestic	.849	.835	.920	.920	.881	.873	.876	.877
Economic	.670	.574	.640	.775	.640	.673	.702	.705
Social control	.057	.090	.040	.052	.099	.074	.069	.050
Civil rights	.001	.009	—	.002	.003	—	—	—
Government	.057	.076	—	.021	.026	.015	.015	.025
Energy	.033	.040	.040	.049	.047	.010	.011	.010
Misc., named	.026	.036	—	.019	.054	.051	.058	.053
Unspecified	.005	.010	.200	.002	.012	.051	.021	.034
Not classified	—	—	—	—	—	—	—	—
Don't know	.016	.031	.040	.024	.023	.017	.021	.018
N cases	1602	1593	1556	1609	1519	1508	1484	1543
N responses	1870	1844	N.A.	1705	1820	1719	1773	1984
Responses/cases	1.17	1.16	—	1.06	1.20	1.14	1.19	1.29

Survey	AIPO1096	AIPO1100	AIPO1103	AIPO1112	AIPO1118	AIPO1125	AIPO1127	AIPO1131
Date	(6/82)	(8/82)	(10/82)	(4/83)	(7/83)	(10/83)	(11/83)	(2/84)
Foreign affairs	.116	.072	.063	.138	.160	.218	.342	.242
Domestic	.869	.906	.922	.842	.814	.749	.629	.729
Economic	.734	.783	.802	.715	.656	.583	.483	.502
Social control	.053	.054	.051	.055	.062	.089	.071	.097
Civil rights	—	—	—	—	—	—	—	—
Government	.022	.016	.018	.017	.018	.024	.024	.019
Energy	.007	.003	.006	.006	.005	.003	.007	.006
Misc., named	.037	.032	.027	.027	.040	.034	.031	.068
Unspecified	.016	.018	.018	.022	.033	.016	.013	.037
Not classified	—	—	—	—	—	—	—	—
Don't know	.015	.021	.015	.019	.026	.033	.028	.028
N cases	1504	1543	1543	1509	1567	1513	1504	1610
N responses	1899	2008	1923	1791	1795	1661	1628	1817
Responses/cases	1.26	1.30	1.25	1.19	1.15	1.10	1.08	1.13

[a] N.A. = not available.

[b] Laos.

[c] Not available from Roper Center; apparently restricted data.

[d] Marginals for AIPO764 in George Gallup, *The Gallup Poll-Public Opinion, 1935–1971* (New York: Random [illegible]use, 1972), III, p. 2151 are actually for AIPO765.

[e] Combines first and second mentions.

A0610-1

POPQAE 49(3) 285–442 (1985)
ISSN 0033–362X

Public Opinion Quarterly

INCOME AND TAXES
  How Much Income Is Enough? *Steven Dubnoff*
  Ideology and the Tax Revolt *Lance deHaven-Smith*
  Public Attitudes on Taxing and Spending *Susan Welch*
Prayer in Public Schools: When Church and State Collide *Kirk W. Elifson and C. Kirk Hadaway*
The Power of *The Right Stuff* *William C. Adams and others*
Viewer Characteristics and Agenda Setting by Television News *David B. Hill*
Finding Subgroups for Surveys *Roger Tourangeau and A. Wade Smith*
Matching Survey Responses to Official Records: Validity in Victimization Reporting *Peter V. Miller and Robert M. Groves*
Explaining Information System Use with System-Monitored vs. Self-Reported Use Measures *James S. Ettema*
Response Scales: Effects of Category Range *Norbert Schwarz and others*
THE POLLS: America's Most Important Problems *Tom W. Smith*

PROCEEDINGS OF THE FORTIETH ANNUAL CONFERENCE OF THE AMERICAN ASSOCIATION FOR PUBLIC OPINION RESEARCH

IN MEMORIAM · NEWS AND NOTES · BOOK REVIEWS

FALL 1985

The Public Opinion Quarterly

VOLUME 49 FALL 1985 NUMBER 3

How Much Income Is Enough? Measuring Public Judgments
Steven Dubnoff 285

Ideology and the Tax Revolt: Florida's Amendment 1
Lance deHaven-Smith 300

The "More for Less" Paradox: Public Attitudes on Taxing and Spending *Susan Welch* 310

Prayer in Public Schools: When Church and State Collide
Kirk W. Elifson and C. Kirk Hadaway 317

The Power of *The Right Stuff:* A Quasi-Experimental Field Test of the Docudrama Hypothesis *William C. Adams and others* 330

Viewer Characteristics and Agenda Setting by Television News
David B. Hill 340

Finding Subgroups for Surveys
Roger Tourangeau and A. Wade Smith 351

Matching Survey Responses to Official Records: An Exploration of Validity in Victimization Reporting
Peter V. Miller and Robert M. Groves 366

Explaining Information System Use with System-Monitored vs. Self-Reported Use Measures *James S. Ettema* 381

Response Scales: Effects of Category Range on Reported Behavior and Comparative Judgments *Norbert Schwarz, Hans-J. Hippler, Brigitte Deutsch, and Fritz Strack* 388

IN MEMORIAM: Archibald Maddock Crossley, 1896–1985
W. Phillips Davison 396

NEWS AND NOTES *Charles M. Oliver, Guest Editor* 398

THE POLLS: America's Most Important Problems
Part II: Regional, Community, and Personal *Tom W. Smith* 403

PROCEEDINGS OF THE FORTIETH ANNUAL AAPOR CONFERENCE 411

BOOK REVIEWS *Marc B. Glassman, Editor* 434
Marilyn Lashner, *The Chilling Effect in TV News: Intimidation by the Nixon White House* *Mark R. Levy* 434

BOOK NOTES *Phyllis M. Endreny, Editor* 436

NOTICE TO CONTRIBUTORS

Three copies of each article should be submitted. Authors are urged to submit original typescripts. Copies of manuscripts which are reviewed become the property of the journal, and unused copies will be returned only if the author provides a stamped, self-addressed envelope. Bibliographic references, footnotes, and tables should be typed to conform to the printed style of this journal.

All submissions that make use of survey data not archieved at the Inter-University consortium for Political and Social Research or the Roper Center must contain the following information: (1) a definition of the population from which the sample was drawn and a statement of the selection procedure sufficient to permit at least approximate replication. (2) The response rate and details of its calculation. (If response rate is not appropriate, the number of refusals.) (3) Dates the survey was conducted. (4) Exact wording of all questions used in the analysis.

Upon acceptance of an article by the journal, the author(s) will be asked to transfer copyright in the article to the publisher. This transfer will insure the widest possible dissemination of information under the U.S. copyright Law.

The PUBLIC OPINION QUARTERLY is hospitable to all points of view, provided that the material presented will help to illuminate problems of communication and public opinion. The American Association for Public Opinion Research assumes responsibility only for material signed by an officer of the Association in his official capacity. Statements of fact and opinion are made on the responsibility of the authors alone and do not represent the views of the editors.

The PUBLIC OPINION QUARTERLY has been registered with the Copyright Clearance Center, Inc. consent is given for copying of articles for personal or internal use, or for the personal or internal use of specific clients. This consent is given on the condition that the copier pay through the Center the per-copy fee stated in the code on the first page of each article for copying beyond that permitted by the U.S. Copyright Law. If no code appears on an article, the author has not given broad consent to copy and permission to copy must be obtained directly from the author. All articles published prior to 1978 may be copied for a per-copy fee of $1.00, also payable through the Center. The consent does not extend to other kinds of copying, such as for general distribution, resale, advertising and promotional purposes, or for creating new collective works.

How Much Income Is Enough? Measuring Public Judgments

STEVEN DUBNOFF

HOW MUCH income is enough? If we are to make sense of this question, we must immediately ask at least three subsidiary questions. First, enough to do what? That is, how much income is enough to achieve a particular standard of living or level of welfare, for example, to "get along" or "not to be poor." Second, enough for whom? People differ in the circumstances that affect their needs for income, such as their family size and retirement status. Finally, enough according to whom? Are we to take the judgments of experts or those of ordinary people considering their own needs or the needs of others? The answers to these questions have important implications for the design and evaluation of income-maintenance and transfer programs and for understanding the mechanism by which objective circumstances, such as income, are translated to subjective well-being.

This article considers these questions and the measurement traditions that are associated with their answers, and presents the results of an

Abstract Respondents were randomly assigned descriptions of families with varying incomes and other characteristics and asked to judge the standard of living of these stimulus families. There is considerable agreement in these judgments. The method yields the income level which is associated with each standard of living and an equivalence scale for adjusting these amounts for differences in household characteristics. The results indicate that the public's judgment of the poverty line is very close to the official level, although the equivalence scales produced by this and other subjective methods imply that the costs of additional family members are much less than those suggested by official scales.

Steven Dubnoff is a Senior Research Fellow, Center for Survey Research, University of Massachusetts/Boston. This article is based on a paper presented at the Annual Meeting of the American Sociological Association, San Antonio, Texas, August 1984. This research was supported by a grant from the Measurement Methods and Data Resources Section of the National Science Foundation. The author would like to thank Miriam Kadin Campbell, Arie Kapteyn, Marian Lowe, Jack Fowler, John Strate, and Denny Vaughan variously for advice, assistance, and moral support.

Public Opinion Quarterly Vol. 49:285–299 © by the Trustees of Columbia University
Published by Elsevier Science Publishing Co., Inc. 0033-362X/85/0049-285/$2.50

experiment in which a sample of individuals was asked to rate the level of living of hypothetical families that differed in income and other characteristics. This experiment yielded evidence of considerable consensus about the income amounts associated with various levels of living and produced results that were comparable with more conventional methods.

Income Minima

There is a long tradition, dating back to at least the beginning of the century, of constructing budgets which purport to measure the amount of income necessary to achieve a given standard of living (Ornatti, 1966). These budgets usually start with the costs of a nutritionally adequate diet and attempt to arrive at a total dollar figure by stating and pricing the components of consumption associated with a given standard of living. The recently abandoned budgets of the Bureau of Labor Statistics and the Federal Poverty line are the well-known models of this tradition.

These standards have been widely criticized because of the necessarily arbitrary judgments that go into their construction and because they rely on the judgments of highly paid experts to determine what is necessary for poorer people to consume and thus involve interpersonal comparisons of utility. Most critically, however, these budgets are constructed in a circular fashion—prevailing patterns of consumption—what *is*—are used to determine normative standards—what is necessary. In fact, the real dollar amounts of such budgets grow fairly directly over time with levels of real consumption or income (Kilpatrick, 1972; Rainwater, 1974; Rein, 1968).

Because of concerns such as these, a committee that recently evaluated the Bureau of Labor Statistics Family Budgets concluded that such quantity budgets should be abandoned as means for estimating the adequacy of incomes. The committee recommended instead that explicitly relative standards should be used as an interim measure, and that, over time, efforts should be directed towards the refinement and validation of techniques to measure popular consumption norms directly (Expert Committee, 1980).

There are several measurement traditions that attempt to bypass expert judgment by using the machinery of survey research to measure popular conceptions of necessary incomes. These techniques grant that norms about consumption are relative to a given time and place and attempt to measure them directly by asking a sample of ordinary people about how much income is thought to be enough. The most venerable survey question in this tradition is the Gallup item, "How much income is necessary for a family of four to get along in this community?" This method has been extended by Rainwater (1974) to different levels of living and family sizes.

Such "other family" minimum income questions ask respondents to suggest income amounts for families in circumstances about which they may or may not have any certain knowledge. A single individual may know nothing of the costs of children, and those with children may know little about the costs of single life. Recently, however, a group of Dutch economists has proposed that respondents name the amount of income that is necessary for them, explicitly considering their own circumstances, to achieve a given standard of living, such as not being poor (Goedhart et al., 1977). The responses to these questions can be analyzed by regression techniques to yield the levels of income at which respondents in various circumstances name their own income, on average, as the minimum income amount. These points can be taken as *the* minimum income.

This article presents a new technique that is a variant on the hypothetical family technique developed by Rainwater (1974). The method involves presenting respondents with randomly assigned families with varying income and other characteristics and asking respondents to rate the standard of living of these stimulus families. In contrast to questions which ask people to name income amounts for various families, a task with which most respondents are probably unfamiliar, this technique asks them to assess how well other families with stated incomes are living. This is more likely to be a task which individuals perform as a part of their everyday lives—we commonly make judgments about who is rich, poor, or just getting along. The following will quantify the factors which go into such assessments and also estimate popular conceptions of the amount of income necessary for those in different circumstances to achieve various standards of living.

Income Equivalence

Given the amount of income necessary for some standard family, for instance, a family of four, to achieve a given standard of living, the next task, in measuring income needs, is to adjust this amount for differences in circumstances. Not to do so is to assume implicitly that two can, in fact, live as cheaply as one. The Federal Poverty line, for instance, is adjusted largely on the basis of variations in the costs of an emergency food plan for families of various sizes and compositions (Orshansky, 1968). Given that the economies of scale associated with food consumption are almost certainly less than the economies of scale associated with total consumption, this technique overestimates the needs of larger families and underestimates the needs of smaller ones (see Dubnoff, 1982).

The problem of adjusting minimum income amounts for variations in family size and other circumstances is identical to the problem of constructing what economists call a "true cost of living index"—the ratio of income necessary for those in differing circumstances to achieve the same level of welfare or standard of living (Kapteyn and van Praag, 1976). For instance, single persons living alone might require half as much income as a family of four to feel equally "well off." If that were true we could divide the income of single individuals by .5 in order to compare their incomes to an income standard for a family of four.

The various subjective techniques, such as those which ask how much income is necessary for respondents to "get along," provide direct and intuitively appealing solutions to the equivalence scale problem. The technique to be described here provides a slightly different ratio: instead of yielding the ratio of incomes necessary for those in different circumstances to *judge themselves* as having the same standard of living, it measures the ratio of incomes necessary for *others to judge them* as having the same standard of living. However, shown below, the results are similar, in most cases, to those derived from techniques which tap respondents' perceptions of their own needs.

Data and Methods

The basic method is to present a sample of respondents with a set of (hypothetical) stimulus families of differing characteristics and have the respondents judge the level of living of each family according to a scale with the following points: Poor, Nearly Poor, Just Getting Along, Living Reasonably Comfortably, Living Very Comfortably, and Prosperous.

THE EXPERIMENTAL DESIGN

The goal of the experimental design was to provide respondents with a reasonably small set of stimulus families, the characteristics of which could be randomly varied along the dimensions which are commonly thought to influence judgments of living levels such as income, family size and composition, age of oldest child, retirement status and sex of head. The income levels were set in a policy-relevant range between 25 percent and 160 percent of the approximate median family income of $21,000.

Although it would be desirable for analytic purposes to produce a balanced design in which the relevant stimulus factors were freely varied across their ranges, this procedure would have produced such improbable families as retired couples with three children under the age of six, or families headed by very young individuals which included eighteen-year-old children. Age was thus brought into the design for families without children by indicating retirement status or a vague yet plausible age ("in his/her/their thirties"). For families with children, the age of the oldest was used to provide a cue about the age of the head.

The basic design included 11 distinct family types, including single-person households headed by retired and nonretired males and females, female- and male-headed households with children, and retired and nonretired couples

Twelve sets of 10 stimulus families were constructed by sequentially assigning the 11 family types, 10 income levels, and, where relevant, the 3 ages of children. The stimulus families were randomly ordered within each set, and the sets of families were randomly assigned to respondents.

THE OUTCOME VARIABLE

The Living Level Scale, which is used as the outcome variable in this analysis, was designed, following pretesting and pilot work, with the goal of producing an ordered scale of commonly understood descriptors of living levels. The scale was intended to provide a reasonable distribution for self-assessments as well as a tool for rating the living levels of others, a task for which more commonly used measures of satisfaction are particularly unsuitable. While the outcome variable certainly has ordinal properties, it was not clear at the outset whether it would or even should have interval properties. We thus asked our respondents to assign a number between zero and 100 to each of the adjectives. The results of this exercise indicated that the points of this scale could, in fact, be treated as if they had equal interval properties.

DATA

The data upon which this analysis is based were collected as part of a much larger effort to evaluate alternative methods for the measurement of income adequacy and equivalence. As such, they were collected in the context of an extensive questionnaire which elicited information on income, income satisfaction, and the respondent's income needs.

The sample was a block quota design within the cities and towns of the Boston SMSA. Given the larger methodological purposes of this study and budgetary constraints, sample eligibility was restricted to households consisting of single individuals or married couples with or without children. The respondent's marital status had to be stable during the previous year. These restrictions serve to eliminate noise in our analysis. Coupled, as they are, however, with the inherent deficiencies of a quota sampling design, they make generalizations to a larger population impossible. However, for the purposes of this analysis, in which experimental conditions are randomly assigned to respondents, these considerations are far less salient.

The data were collected by trained, professional interviewers in the period between April and June, 1983. Altogether, 588 interviews were conducted. With 10 experimental questions per respondent and allowances for missing data, 5590 responses are available for analysis. Although we will report sampling errors computed on the assumptions of simple random sampling, these are smaller than they would be if allowances were made for the clustering inherent in the sample design and the fact that 10 hypothetical families were rated by each respondent.

Relationship of Responses to Stimulus Family Income

Table 1 indicates a strong relationship of stimulus family income to the living level responses. It shows that when our respondents assessed the living level of these hypothetical families, they quite clearly and overwhelmingly took the income of the stimulus family into account. The responses are clustered on the diagonal, and there is a high association (gamma=.72) between stimulus family income and the responses.

Table 1. Responses for Stimulus Family Standard of Living by Stimulus Family Income (Percent)[a]

Standard of Living	*Family Income (in thousands of dollars)*										*All Income Groups*
	4.20	*7.35*	*9.45*	*12.6*	*15.75*	*17.85*	*21.0*	*24.15*	*27.3*	*32.5*	
Poor	83	56	40	19	8	6	3	1	1	0	22
Nearly poor	12	26	31	33	24	17	9	6	3	1	16
Just getting along	4	15	21	35	43	45	40	27	20	8	26
Living reasonably comfortably	1	3	6	11	20	25	37	45	49	41	24
Living very comfortably	1	1	1	2	5	6	10	18	23	40	11
Prosperous	0	0	0	0	1	0	1	3	4	10	2
Total	100	100	100	100	100	100	100	100	100	100	100

[a] Rounding may affect column totals.
Goodman-Kruskal Gamma 0.72
Kendall's Tau-B 0.62
N=5772

Figure 1, which is based on the data in Table 1, displays the plot of the percentage of responses in four living level categories against the income of the stimulus family. The percentage of families that are judged to be poor drops precipitously from a high of over 80 percent at an income of about $4,000 to below 10 percent at incomes about $16,000. The distributions for Nearly Poor and Getting Along are nearly normal, with peaks,

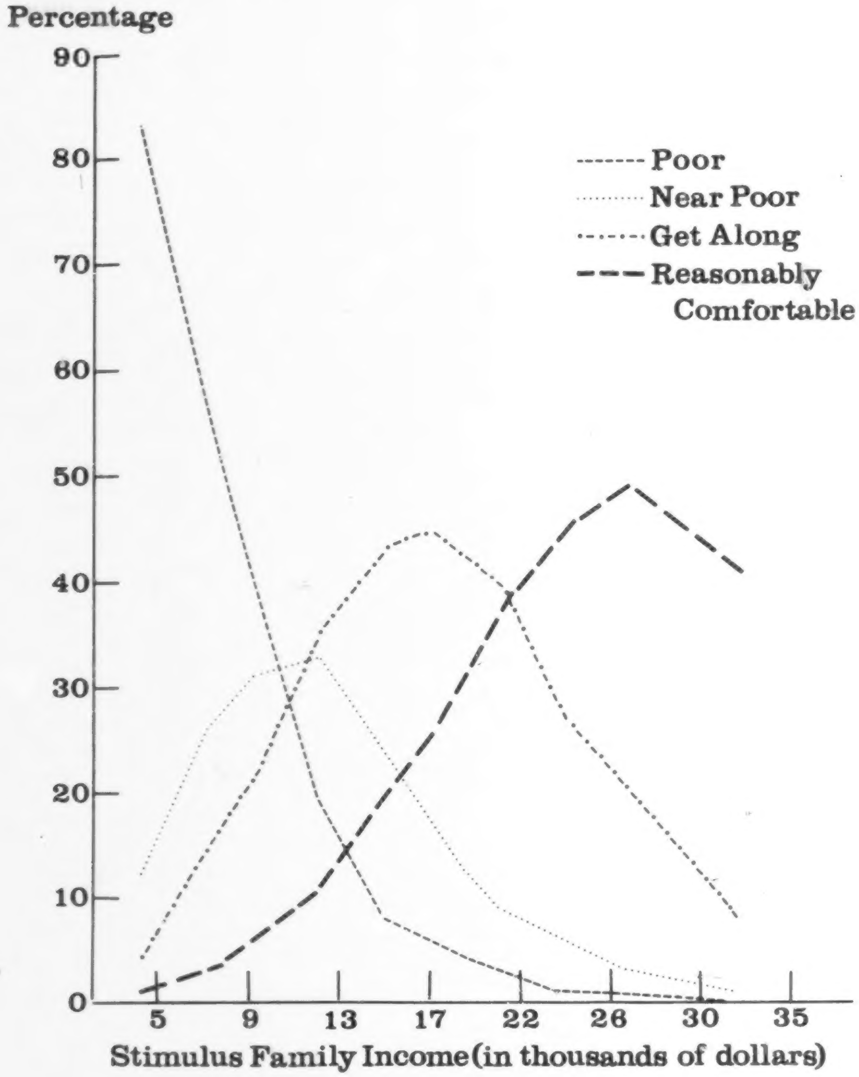


Figure 1. Living Level Judgments

respectively, at about $12,000 and $17,000. Living Reasonably Comfortably peaks at about $27,000, and the right tail of its distribution is outside the range of the experimental income conditions. The responses for Living Very Comfortably and Prosperous, which are not graphed, would show that the range of experimental income conditions was not high enough to capture their peaks.

There is thus a very clean zero-order relationship between the living level judgments and the stimulus families' income. However, the stimulus families varied in size from one to six members within each income group.

Therefore, more powerful analytic techniques are required to control for this variation in the situations of the stimulus families as well as for the effects, if any, of respondent characteristics on the data. We will turn to this in the following section.

Determinant of Living Level Judgments

We will discuss two models of the determinants of respondents' assessments of hypothetical family living levels. These are presented in Table 2. The first includes the characteristics of the stimulus families, and the second adds the characteristics of the respondents making the judgments.

The first model consists of the log of income, the log of family size, and the retirement status and sex of the head of the hypothetical family. These four variables explain almost 61 percent of the variance of the living level scales. Because the stimulus family characteristics are uncorrelated by design, an inspection of the marginal R^2 for each variable can provide some insight into the importance of each factor. First, it is clear that the overwhelmingly largest part of the explained variance is due to

Table 2. Effect of Stimulus Family and Respondent Characteristics on the Evaluation of the Standard of Living of Hypothetical Families

Independent Variables: Stimulus Family and Respondent Characteristics	*Model 1*		*Model 2*	
	Beta (S.E.)	*Marginal R^2*	*Beta (S.E.)*	*Marginal R^2*
Stimulus family				
Ln (income)	1.664	0.546	1.664	0.545
	(0.019)		(0.018)	
Ln (family size)	−0.547	0.036	−0.548	0.036
	(0.024)		(0.023)	
Retirement status	0.223	0.004	0.221	0.004
(1=retired)	(0.029)		(0.028)	
Sex	0.054	0.001	0.055	0.001
(1=female)	(0.013)		(0.012)	
Respondent				
Ln (income)			−0.312	0.027
			(0.015)	
Ln (family size)			0.146	0.003
			(0.021)	
Retirement status			0.288	0.005
(1=retired)			(0.033)	
Sex			0.015	0.000
(1=female)			(0.022)	
Constant	−12.789		−9.832	
	(0.182)		(0.228)	
R^2	0.6084		0.6458	

$N = 5590$

the income of the stimulus family. The size of the family is next in importance. The log of family size accounts for about 3.5 percent of the variance. Larger families, compared to smaller families with the same income, are judged to have a lower standard of living. Both the retirement status and sex of the head have relatively small but statistically significant effects on the living level judgments. Households headed by retired individuals or female-headed households are judged to be better off, controlling for income and family size, than other households.

The second model includes respondent characteristics as well as stimulus family characteristics. Because these two sets of variables are uncorrelated with each other, the coefficients and contributions to R^2 of the stimulus family characteristics are nearly identical in the first and second model. Taken together, the log of respondent's income, log family size, retirement status and sex add about 4 percent to the explained variance.

Respondent's income adds about 2.5 percent to the explained variance. This means that the respondents' vantage point influences their judgments. The negative sign on this variable indicates that respondents with higher incomes tend to assign lower standards of living than respondents with lower incomes. In addition, respondents' family size and retirement status have significant effects. Respondents in larger families and retired respondents tend to rate the standard of living of the stimulus families higher. These two variables together account for about 0.8 percent of the explained variance. Respondent's sex has no effect on the responses.

The most notable aspect of the regression results is the large amount of variance explained by the income of the stimulus family. Although there is an effect of respondent's own income on these judgments, the effect is quite small relative to the effect of the stimulus family's income. Finally, it is worth noting that roughly 60 percent of the variance is explained by a model which includes only the demographic characteristics of the stimulus family. This is an extraordinarily good fit for a nontrivial social science microdata model in which an attitude is used as an outcome variable. It is thus fair to state that a sample of ordinary people exhibit considerable agreement when making judgments about the living level of others—that there is consensus about the amount of income which is necessary to reach various standards of living.

How Much Income Is Enough

The regression results presented in the preceding section can be used to estimate both income equivalence scales and the income levels which are associated with the various points on the living level scale. Recall that we have fit a regression model of the relationship between stimulus family characteristics and the living level judgments. We can use the estimated coefficients of this regression model to solve for income in terms of the other variables. Given this transformed equation we can substitute values of the living level scale and hypothetical family characteristics and solve for the income levels associated with those characteristics. These computed income amounts can be taken as the levels of income which are associated, on average, with each living level for each family type. The ratios of these amounts are an equivalence scale, the ratio of income such that families of different types are judged to have the same standard of living.

The results of this analysis are presented in Table 3, which presents the estimated income amounts for households of one to six persons headed by a nonretired individual, and retired people in one- and two-person households. Consider the column for the standard family of four. Recall that the outcome variable is a six-point scale coded one for Poor and six for Prosperous. The interval between Poor and Nearly Poor is thus in the range from one to two. The logical dividing line between Nearly Poor and Poor is thus 1.5, an amount which is just under $9,000. The midpoint of the interval between Nearly Poor and Getting Along is 2.5, which corresponds to an income of about $15,800. The amount at 3.0, which can be taken as the center of Getting Along space is about $21,000, or about the median income. Similarly, the dividing line between Getting Along and Reasonably Comfortable is about at $29,000.

Table 3. Standard Income in Dollars and Equivalence Scale for Various Living Standards by Retirement Status and Family Size

Standard of Living	*Family Size*							
	Nonretired						*Retired*	
	1	*2*	*3*	*4*	*5*	*6*	*1*	*2*
1.0 Poor	$3,971	$5,048	$5,808	$6,417	$6,932	$7,384	$3,473	$4,414
1.5	5,362	6,817	7,844	8,666	9,362	9,972	4,690	5,962
2.0 Near poor	7,242	9,206	10,594	11,703	12,643	13,467	6,333	8,051
2.5	9,780	12,433	14,307	15,805	17,075	18,188	8,553	10,874
3.0 Just getting along	13,208	16,791	19,322	21,345	23,060	24,563	11,551	14,685
3.5	17,838	22,676	26,094	28,827	31,143	33,172	15,600	19,832
4.0 Living reasonably comfortably	24,090	30,624	35,240	38,931	42,058	44,799	21,068	26,783
4.5	32,533	41,358	47,592	52,577	56,800	60,501	28,453	36,171
5.0 Living very comfortably	43,937	55,855	64,273	71,005	76,709	81,708	38,426	48,849
5.5	59,337	75,432	86,802	95,893	103,596	110,347	51,895	65,971
6.0 Prosperous	80,135	101,872	117,226	129,504	139,907	149,024	70,084	89,095
Equivalence scale	62	79	91	100	108	115	54	69

As noted above, the range of experimental income amounts was, by design, not high enough to capture the upper parts of the scale. We are thus on fairly uncertain ground when estimating the amounts which correspond to higher living levels. In any event, the amount which corresponds to the lower bound of Living Very Comfortably (4.5) is about $52,600 and of Prosperous (5.5) is about $96,000.

The equivalence scale, which is presented in the last row of Table 3, does not depend on the living level which is chosen. This scale implies that one-person households require about 60 percent of the income of four-person households to achieve the same standard of living. Retired families require slightly less income than nonretired families.

A Comparison to Other Methods

The foregoing has presented a new method for measuring popular conceptions of the amount of family income in different situations which correspond to various levels of living. As discussed at the outset, there are a number of other methods for arriving at the same sort of series. What follows puts the results of this exercise in context by comparing the results to those derived from various other methods.

INCOME MINIMA

Table 4 presents income minima which are derived by other methods for two levels of living, the poverty line and Getting Along. The amounts are presented for two family types, a "standard" family of four and a retired couple.

Looking first at the poverty level, we see that the income associated with 1.5, or the midpoint between Poor and Nearly Poor, is within $1,000 of the Federal Poverty Line for retired couples and families of four. It is about $1,600 lower than the results derived from a minimum income question which asked respondents to name the largest amount of income which they needed, considering their own circumstances, in order to "not be poor." Finally, the results from the hypothetical family method presented in this article for the levels 1.5 and 2.0 bracket these other methods. These results indicate that the poverty line can be fairly well located with this method.

For Getting Along, the results at 2.5, the boundary between Getting Along and Nearly Poor is lower than the other measures. For a family of four, the results are almost $5,000 below those from the minimum income question and about $3,000 below those of the Bureau of Labor Statistics Lower Budget. For retired couples, the results presented in this article are more in line with these other standards. Finally, a question in the Gallup/Rainwater tradition, which asks respondents how much income

Table 4. Comparison of Minimum Income Amounts by Method for Families of 4 and Retired Couples

Method	*Family of 4*	*Retired Couple*
Poverty level		
This article (1.5)	$8,666	$5,962
Federal Poverty Line (1982)	9,316	5,413
Minimum income question[a]	10,256	5,413
This article (Near Poor, 2.0)	11,703	8,051
Getting along		
This article (2.5)	15,808	10,874
BLS lower budget[b]	18,639	8,789
Minimum income question[a]	20,769	11,813
This article (3.0)	21,345	14,685
Other family dollar needs[c]	22,101	15,563
BLS intermediate budget[b]	30,904	12,474

[a] Computed by author. Same dataset as this article. MinY according to the Dutch method (Goedhart, et al., 1977). Questions refer to largest amount of income at which respndent would "not be poor" and smallest amount necessary to "just get along."

[b] Autumn 1981 budgets (U.S. Bureau of Labor Statistics 1982a, 1982b) were multiplied by a correction factor of 1.21673, to compensate for the Boston index (115) and the change in the CPI to the survey period.

[c] Computed by author. Same dataset as this article. Geometric mean of responses to questions about the dollar needs of the two family types to Get Along.

is necessary for a hypothetical family of four to "get along," produces results which are higher than any of the other methods.

INCOME EQUIVALENCE

Table 5 presents a comparison of Income Equivalence Scales by the method used in their construction. First, considering two methods in which respondents are asked to make judgments about randomly assigned hypothetical families, it can be seen that answers are very similar, whether respondents are asked to judge the living level of families with given income and other characteristics, as in this article, or to name the income amounts which are necessary for families with given standards of living and other characteristics. The results of both of these methods are in line with those derived from the minimum income question, in which respondents are asked to name the income necessary for them, considering their own circumstances, to not be poor. The scales produced for the same question with a "getting along" reference point are, however, slightly steeper.

When we consider the results for retired people, an interesting difference emerges. When asked about their own needs, retired people tend to think that they need considerably less income than nonretired individuals. When others judge the needs of retired people, as in row one of Table 5, the difference is less marked. In other words, retired individuals think

Table 5. Comparison of Income Equivalence Scales by Method by Retirement Status and Family Size

	Family Size (N Persons)							
	Nonretired						Retired	
Method	1	2	3	4	5	6	1	2
Hypothetical family (standard of living)[a]	62	78	90	100	108	115	54	69
Hypothetical family (income)[b]	—	80	91	100	107	114	—	—
Minimum income (poor)[c]	64	80	91	100	108	114	44	57
Minimum income (get along)[d]	58	76	90	100	109	117	44	57
Poverty line[e]	50	65	78	100	118	134	47	59
BLS[f]	39	61	83	100	124	148	28	52

[a] This article, hypothetical families, varied income and characteristics. Respondents rate standard of living. Scale is income ratio for families in different circumstances to be judged as having the same standard of living.

[b] Rainwater (1974:102,105). Respondents judged income amounts necessary for hypothetical families to achieve given standards of living. Ratio of natural logs of responses, by family size.

[c] Computed by author, Boston SMSA data. Responses to question about the highest income amount at which they would not be poor.

[d] Computed by author, Boston SMSA data. Responses to question about lowest income amount they would need to "just get along."

[e] Federal Poverty Line (Orshansky, 1965). Based on ratios of costs of Economy Food Plan, except for sizes one and two.

[f] Bureau of Labor Statistics Equivalence Scale (Jackson, 1958), constructed to equalize proportion of total expenditures devoted to food by family type.

they need less income than the public at large thinks they need. This has important implications for policy in our period, in which a shrinking pool of working people support the needs of those who are retired through various transfer programs.

Finally, the scales which are based on food consumption or needs, most particularly the Bureau of Labor Statistics Scale, are considerable steeper than any of the subjective, survey-based scales, and in addition imply that retired or older individuals need considerably less income than those who are not aged or retired.

Discussion

The foregoing exposition has demonstrated that there is considerable agreement about how much income is enough for families in various circumstances to reach various levels of living. With respect to the needs of others, people agree about how much income is enough, but that does not mean that they agree about their own needs. We can explain roughly 60 percent of the variance in the judgments of the living level of others with a few demographic characteristics; when we asked our respondents about their own level of living, however, a much more complicated model is needed to explain much less variance. There thus exists a set of general norms about the correspondence between income and standard of living, within which individuals can and do make judgments about their own well-being that differ from the judgments they make about others when represented with just a few demographic facts.

The information that was presented in this experiment is, however, close to the data which are relevant in the context of income-maintenance policy. The large degree of consensus displayed by our sample supports the notion that the norms which tap income needs are both salient and easily measurable. Samples of ordinary people can thus be the real "experts" on how much income is enough.

The results presented here indicate that the precise method used to measure opinions about income norms, particularly with respect to estimating income equivalence, does not make much difference. Research is now underway which will help to establish which method is to be preferred, but, whatever the outcome, the preponderance of evidence indicates that survey-based measures can provide useful and policy-relevant answers to questions about how much income is enough.

References

Dubnoff, Steven
1982 "A method for the validation of equivalence scales." Paper presented at the annual meetings of the American Association of Public Opinion Research.
Expert Committee for Family Budget Revisions
1980 New American Family Budget Standards. Madison, WI and New York: Institute for Research of Poverty and the Center for Social Sciences, Columbia University.
Goedhart, T.V., V. Halberstadt, A. Kapteyn, and B.M.S. van Praag
1977 "The poverty line: concept and measurement." Journal of Human Resources 12:503–520.
Jackson, Carolyn A.
1958 "Revised equivalence scales for estimating income or budget costs by family type." U.S. Bureau of Labor Statistics, Bulletin 1570-2.
Kapteyn, Arie, and Bernard van Praag
1976 "A new approach to the construction of equivalence scales." European Economic Review 7:313–35.
Kilpatrick, R.W.
1972 "The income elasticity of the poverty line." Review of Economics and Statistics 55:327–332.
Rainwater, Lee
1974 What Money Buys: Inequality and the Social Meanings of Income. New York: Basic Books.
Rein, Martin
1968 "Problems in the definition and measurement of poverty," in L. Ferman (ed.), Poverty in America. Ann Arbor: University of Michigan Press.

Ornatti, Oscar
1966 Poverty Amid Affluence. New York: The Twentieth Century Fund.
Orshansky, Mollie
1958 "Counting the poor: another look at the poverty profile." Social Security Bulletin 18:3–19.
U.S. Bureau of Labor Statistics
1982a "Family budgets. Final report on family budgets." Monthly Labor Review 105(7): 44–46.
U.S. Bureau of Labor Statistics
1982b "Family budgets. Retired couple's budgets." Monthly Labor Review 105(11): 37–38.

The "More for Less" Paradox: Public Attitudes on Taxing and Spending

SUSAN WELCH

ONE OF THE ostensible paradoxes of public attitudes in recent years has been the simultaneous desire of many citizens for both tax reductions and increased levels of governmental services. Sears and Citrin (1982), for example, show that well over 70 percent of Californians before and after Proposition 13 (1978) thought their taxes were too high, yet majorities preferred increases in spending for a wide variety of services—mental health, police, fire, prisons and corrections, schools, and transportation. Only for welfare did a majority of the public favor cutbacks (support for increased spending has been found by Wilson, 1983; Citrin, 1979; Ladd, 1979; Courant et al., 1980; Arrington and Jordan, 1982). Sears and Citrin (1982:54–55) also reported that even those who believed that government is too big and their own taxes much too high favored increased spending.

On its face, these public preferences for tax reductions and more services seem wildly contradictory (but see Musgrave, 1979; Lowery and Sigelman, 1981). To hold these two views may indeed be irrational, but it is also possible that citizens, just like legislators, executives, and bureau-

Abstract Using a statewide random sample of over 900 respondents, attitudes toward taxing and spending are examined in order to try to explain the seeming paradox of a public that wants more spending but less taxation. We investigate the possibility that the public which wants more spending is willing to pay for it by means other than taxes, such as through reallocation or increased nontax revenue. Moderate support for this expectation is found. Overall, the desire for a "free lunch" is not as widespread as a simple comparison of taxing and spending preferences suggests.

Susan Welch is Carl A. Happold Professor of Political Science, University of Nebraska-Lincoln. The author gratefully acknowledges the financial support of the University of Nebraska Research Council for this project. John Hibbing made many helpful comments on the paper.

Public Opinion Quarterly Vol. 49:310–316 © by the Trustees of Columbia University
Published by Elsevier Science Publishing Co., Inc. 0033-362X/85/0049-310/$2.50

crats, hold preferences for increased funding that is not financed by taxes. President Reagan and Congress have financed increased military spending through borrowing. States and localities normally have strict borrowing limits but have increased user fees as revenue sources.

Beginning to chart these heretofore unexplored preferences is the object of this article. For each of the three levels of government, we will explore several alternative possible public preferences for financing increased spending other than increased taxes.

Data and Methods

Data from a spring (February–April) 1983 sample of households in one midwestern state are the basis of our study. Random digit dialing was used to obtain a random sample of approximately 900 households, and within each household one adult was randomly selected for inclusion in the study. A response rate of 84 percent was obtained. The final sample matches well the aggregate characteristics of the adult population of this state, except that women were overrepresented. A weighting factor was applied so that the gender division of the sample would approximate that of the state population.

Respondents were asked whether or not they wished to spend more on each of eight areas: energy, health care, the elderly, the poor and unemployed, farm problems, education, crime prevention, and water resources. Separate responses were recorded for each level of government.

Those respondents who wanted to spend more in at least one area (and more than 90 percent did) at each level of government were then asked:

How should state government pay for doing more in the area(s) of (name areas where respondent wanted more spending):
  a. Increase state sales or income tax.
  b. Practice greater efficiencies in state government programs.
  c. Do less in other state programs.
  d. Increase aid from federal government.
Same item as above, substituting "federal" for "state."
  a. Increase taxes.
  b. Practice greater efficiencies in national government programs.
  c. Spend less on military and defense.
  d. Spend less on other national programs.
Same item, substituting "local" for "state."
  a. Increase sales tax.
  b. Increase local property tax.
  c. Increase state aid to local government.
  d. Increase federal aid to local government.
  e. Practice greater efficiencies in local government.
  f. Do less in other local programs.
  g. Increase fees for services (user fees, licenses, etc.).

Categories of Support for Revenue Raising

From these responses we have devised a fourfold categorization that taps the willingness of individuals to raise revenue to support increased spending. We call this the "willingness to pay" scale. Respondents who wanted to increase spending in at least one area were placed in one of the following categories.

1. *Taxers*—those willing to increase taxes. This seems like the most concrete expression of willingness to pay for new or expanded services. While many previous studies have investigated attitudes toward taxes, only a few have juxtaposed tax and spending preferences. Wilson, for example, found that between 62 percent and 84 percent of respondents wishing to increase local spending for specified services were willing to pay increased taxes (but see Arrington and Jordan, 1982).

2. *Probably realistic revenue raisers.* In this category were placed individuals who were not willing to raise taxes to finance spending but who were willing to do one of the following:

a. *Adopt new user fees.* This indicates a willingness to generate new revenue, but this response was not placed in the same category as support for increased taxes because the revenue likely to be raised may only be appropriate for narrowly defined services (cf. Mushkin and Vehorn, 1980).

b. *Increase revenue* through intergovernmental aid coupled with taxing to raise revenue at the grantor level. This indicates a realistic desire to provide revenue for increased spending but to transfer the locus of the spending to a higher government level, as in increased state aid for education.

c. *Reallocate* coupled with a pattern of responses indicating some realism in that preference. An expressed desire to reallocate may be realistic or less realistic. Wilson (1983) found that between 10 percent and 28 percent of his two-city sample were willing to reallocate to pay for service increases in priority areas. Those who claim they want to reallocate to pay for increased health care benefits, for example, but who also desire more spending in every other area, are probably not being very realistic about reallocation. We put respondents in this category if their expressed preference for reallocation was coupled with a desire to increase spending in not more than half of the functions about which they were asked. At the federal level, respondents were also placed in this category if they wanted to reallocate through decreased military spending regardless of the number of domestic reallocations desired. Obviously, our measure of the realism behind reallocation preferences is far from perfect, but at least it is an attempt to gauge the seriousness of reallocation intent.

3. *Probably not realistic revenue raisers.* Here we put individuals who expressed preferences for reallocation but wanted to spend more in more

than half of the functions. We also placed in this category those who wanted more intergovernmental aid, but were not willing to tax more at the grantor level. These responses seemed of the "something for nothing" type, yet were a bit more concrete than the responses in our final category. It is possible, for example, that respondents really are willing to reallocate drastically even if they did favor increased funding in most service areas, or are willing to reallocate at the grantor level and make increased intergovenrmental aid possible without increased taxes at the grantor level. Nevertheless, these responses seem less realistic than those in category 2.

4. *Free lunchers.* Respondents in this category are not supportive of increased taxation, reallocation, or intergovernmental aid. They want to finance increased spending entirely by greater governmental efficiencies. Indeed, the vast majority of the public believes that government is wasteful. SRC surveys indicate that while in 1964 only 47 percent thought government wasted some money, by 1980 this proportion was 82 percent. Only 2 percent believed the government did not waste any money. In California, before Proposition 13, a majority felt that the same level of services could be maintained with a 10 percent cut in spending, and a near majority felt that even a 20 percent cut would not result in service reduction (Sears and Citrin, 1982:51). This attitude seemed to us the purest "free lunch" response.

Preferences for Funding Increased Services

The upper part of Table 1 shows the proportions favoring each mode of revenue raising, and the lower part shows how the sample fits in our "willingness to pay" scale. Those individuals willing to tax (over a quarter of the sample at the federal level and nearly one-half at the state and local) are placed in the first category of our scale.

At the second level are those whose willingness to pay seems realistic; while not favoring increased taxes, they do evidence a seemingly serious desire to reallocate, a preference for a shift in the locus of funding to a higher government level, or a preference for user fees. The proportion fitting this category is considerably lower for state government (less than one quarter) than for either local or federal ones.

Table 2 illustrates how these proportions are derived. Among those initially eligible for category 2 (that is, those not favoring increased taxes but favoring intergovernmental aid), only a minority are willing to tax at the grantor level to produce increased revenue (A). And, among those stating a willingness to reallocate, between 27 percent and 44 percent at the same time want to spend more on over half of the categories (B). Within the category of federal reallocations some prefer to decrease

Table 1. How the Public Wishes to Pay for Increased Services,[a] and How the Sample Fits in "Willingness to Pay" Scale[b]

	National	*State*	*Local*
Increased services funded by:			
Increased taxes	27%[a]	49%	48%
User fees	—	—	40
Intergovernmental transfers	—	44	77
Reallocation	88	64	61
Efficiency	98	96	96
Fourfold scale of willingness to pay			
Taxers	27%	49%	48%
Realistic revenue raisers[c]	51	23	40
Probably not realistic revenue raisers[d]	14	14	5
Free lunch—not willing to tax, assess user fees, reallocate or increase aid	8	14	7
	100%	100%	100%

[a] Percents add to more than 100 because of multiple responses.

[b] Only those wishing one or more increased services included. $N=809$.

[c] Do not support increased taxes but do support reallocation coupled with preference to increase spending in only 1–4 areas, or at federal level, reduce military spending; or intergovernmental aid preference with willingness to increase taxes at higher level.

[d] Preferences for reallocation or intergovernmental aid without meeting either condition in c.

military spending even though they want increases in most domestic programs (C).

At the third level are placed those respondents whom we have labeled "unrealistic revenue raisers;" they do not favor tax increases and their desire for raising other kinds of revenue is not consistent with their other responses. This category is identical at the federal and state or local levels (14 percent). At the local level, the number in this category is fairly negligible (5 percent), indicating that with regard to local revenue raising, people are somewhat more realistic than they are in thinking about other levels of government.

Finally, the last category contains only a small proportion of people for all levels. These are respondents who wished to increase spending but favored only increased efficiencies in government as the method for doing so. This proportion ranged from 7 percent to 14 percent at the three levels.

On the whole, then, our scale suggests that relatively few want a "free lunch," more spending without any reasonable means to pay for it. While some do not support increased taxes, many are willing to shift the burden of government to a higher level with consequent increased taxes there, or reallocate in a meaningful way, or assess more user charges. From 72

Table 2. Deriving "Willingness to Pay" Scale Positions

A. *Want Intergovernmental Aid but No Taxes*

	Aid to State Government	*Aid to Local Government*	
Willing to tax at grantor level	13%	24%	"Realistic"
Not willing to tax but willing to reallocate at grantor level	74	72	
Neither tax nor reallocate at grantor level	8	4	"Not Realistic"
	100%	100%	
N	(168)	(135)	

B. *Want to Reallocate but Not Tax*

Number of areas of increased spending	*Federal*	*State*	*Local*	
1–2 Areas	22%	26%	38%	"Realistic"
3–4 Areas	34	35	36	
5–6 Areas	29	28	23	"Not Realistic"
7–8 Areas	15	12	4	
	100%	100%	100%	
N	(531)	(231)	(141)	

C. *Federal Reallocation*

Willingness to reallocate only from domestic spending	33%
Willing to reallocate from military spending with or without domestic reallocation	39
Not willing to support reallocation from the military and favors spending more in 5–8 domestic areas	28
	100%

percent to 88 percent of the respondents are found in one of the first two categories, where support for raising revenue in realistic ways is apparent. From 12 percent to 22 percent are found in the other two categories, where spending is desired but means to pay for it are not very concrete.

Implications

This research has tried to expand our awareness of public preferences about raising and spending money by examining the relationship between preferences for increased spending and support for ways to fund that spending. We have found that support for expansion of public services is high. We have shown that the inconsistency of citizens wanting more for less, more services and less spending, probably is a paradox for only a minority. Most citizens are willing to raise additional revenue to pay for these services, or at least to reallocate from less desired to more desired services. Very few citizens think we can rely solely on greater efficiencies in government to pay for increased services, and only 12–22 percent seem to favor reallocation or intergovernmental aid in a way inconsistent with their preferences.

Thus we have found that in the abstract at least, not as many Americans are looking for a "free lunch" as a simple comparison of taxing and spending preferences would suggest. For many, reallocation preferences or shifts in the locus of spending seem real. Still, we do not know whether citizens have very accurate ideas of how much money might be raised by reallocating from their less desired services nor do we know how they would balance cost against benefits in their reallocation preferences. Nevertheless, this research does indicate the taxing-spending paradox may not be as great as previously thought.

References

Arrington, Thomas and David Jordan
1982 "Willingness to Pay Per Capita Costs as a Measure of Support for Urban Services," Public Administration Review 42 (March/April):168–170.
Citrin, Jack
1979 "Do People Want Something for Nothing," National Tax Journal 32 (June):113–129.
Courant, Paul, Edward Gramlich and Daniel Rubenfeld
1980 "Why Voters Support Tax Limitation Amendments: The Michigan Case," National Tax Journal 33 (March):1–20.
Ladd, Everett Carl
1979 "The Polls: Taxing and Spending," Public Opinion Quarterly 43 (Spring):126–135.
Lowery, David and Lee Sigelman
1981 "Understanding the Tax Revolt: Eight Explanations," American Political Science Review 75 (December): 963–974.
Musgrave, Richard
1979 "The Tax Revolt," Social Science Quarterly 59:697–703.
Mushkin, Selma and Charles L. Vehorn
1980 "User Fees and Charges," in Charles Levine, ed. Managing Fiscal Stress. Chatham House, 222–234 Chatham, N.J.
Sears, David and Jack Citrin
1982 Tax Revolt Cambridge: Harvard University Press.
Wilson, L. A.
1983 "Preference Revelation and Public Policy: Making Sense of Citizen Survey Data," Public Administration Review 43 (July/August):335–343.

Prayer in Public Schools: When Church and State Collide

KIRK W. ELIFSON AND C. KIRK HADAWAY

THE UNITED STATES has evolved one of the clearest separations of religion from government; nevertheless, the issue of religion in public schools has generated considerable controversy. Prayer and Bible reading in public schools have led to three major Supreme Court decisions; they are the impetus for approximately 200 constitutional amendments in the U.S. Congress, amendments designed to permit voluntary prayer in public schools or curtail federal court jurisdiction in school prayer cases.

Even with this series of proposed amendments and considerable interest by the mass media, little effort has been given to analyzing the phenomenon. Detailed knowledge concerning the social and demographic characteristics of those who favor prayer in the schools (and those who oppose it) is scarce.[1] This article addresses this issue through an

---

[1] A 1964–65 national survey of public elementary school teachers was reported by Way (1968), and Adams (1975) analyzed 1960s public opinion data on school prayer.

Abstract Prayer and Bible reading in public schools have led to three major Supreme Court decisions and the introduction of numerous constitutional amendments in the U.S. Congress which would permit voluntary prayer in public schools or limit federal court jurisdiction. Public opinion polls beginning in 1964 have suggested widespread support among the American public for prayer in public schools. Little is known, however, concerning the characteristics of those supporting and opposing prayer in public schools. This article addresses the school prayer issue through an analysis of three national surveys which were conducted in 1974, 1980, and 1982. Those favoring school prayer were found to be older, less educated, and socially, politically, and religiously conservative. Multivariate analysis revealed that the key predictor variables were religious orthodoxy and religious salience, a finding which may stem from the observation that school prayer is both a political and a religious issue. Issue salience, congressional voting decisions, and demographic characteristics of school prayer advocates are considered as possible explanations for the failure of Congress to act in accord with public opinion.

Kirk W. Elifson is Associate Professor in the Department of Sociology, Georgia State University, Atlanta, Georgia. C. Kirk Hadaway is Research Director, Center for Urban Church Studies, Nashville, Tennessee. The authors wish to thank Jim Garand, John Hutcheson, and G. Melton Mobley for their helpful comments on earlier drafts.

Public Opinion Quarterly Vol. 49:317–329 © by the Trustees of Columbia University
Published by Elsevier Science Publishing Co., Inc. 0033-362X/85/0049-317/$2.50

analysis of empirical data from three national surveys which assessed respondents' views on prayer in the public schools.

Religion in State-Supported Schools

Ironically, the first schools in the United States were religious and not secular. By the beginning of the Revolutionary War, state churches had been established in a majority of the colonies; not until the 1830s were serious objections offered to Bible reading in the schools. The immigration of large numbers of Roman Catholics to the East Coast in the early nineteenth century followed later by Jews and other religious minorities resulted in the first serious protests concerning religion in the schools. Most states forbade the use of public funds for church schools by the 1870s (Stokes and Pfeffer, 1964).

Religion in the form of Bible reading, devotional exercises, and prayer remained an integral part of many school systems until 1960. At this point three Supreme Court decisions that dealt with prayer in public schools were rendered within a period of one year. All centered upon one sentence in the First Amendment to the United States Constitution which simultaneously guaranteed government neutrality toward religion (the "Establishment" clause) and the right to freely practice religion (the "Free Exercise" clause).[2] The three cases upon which the current controversy rests include *Engel* v. *Vitale* (1962), which tested the constitutionality of a daily school prayer written by the New York State Board of Regents,[3] and *Abington School District* v. *Schempp* (1963), which questioned a Pennsylvania law requiring the reading of 10 Bible verses without comment to begin the school day. The third case, *Murray* v. *Curlett* (1963), which was decided by the *Schempp* case, challenged the practice of Bible reading and recitation of the Lord's prayer in the Baltimore schools. In all three instances, the Court ruled the activities unconstitutional, and the effect was to remove legal expressions of organized prayer and Bible reading from the public school.

Religion and Politics

The question of why people oppose decisions made by the Supreme Court concerning school prayer can be seen as both political and religious. It is political in the sense that the issue concerns public policy enforced by the state, and it is religious in that it concerns the expression of a religious belief or feeling. Political scientists generally view the role of religion vis-à-vis politics as minimal (Lane, 1959), yet in this case the two are so intertwined that an association would seem almost axiomatic. Our expectation, therefore, is that personal religiosity should have a relationship with attitudes about returning prayer to the school which is stronger than the "minimal" association which might otherwise be anticipated when investigating religion and politics.

---

[2] "Congress shall make no law respecting an establishment of religion or prohibiting the free exercise thereof." (U.S. Constitution, Amendment I).

[3] "Almighty God, we acknowledge our dependence upon Thee and we beg Thy blessings upon us, our parents, our teachers and our country."

It is also necessary when considering this issue to look at various motivations which may work in concert with or in opposition to religiosity. With such a large number of persons approving the return of prayer to the public schools, it is clear that people with a wide variety of social characteristics share this attitude.[4] Thus, while no isolated, peculiarly distinctive subgroup favoring prayer in schools exists, certain social characteristics which predispose many individuals towards acceptance of school prayer can be illuminated. The first clue comes from those who are most vocal in pushing for constitutional amendments: southern evangelicals, a group who Perkins et al. (1983) report favor allowing prayer in public schools by a margin of 94 to 6 percent. Accordingly, social and political conservatism, region of residence, and denominational affiliation are key variables. Finally, persons who are generally less tolerant, are tradition oriented, or are localistic and parochial in world views are expected to oppose recent Supreme Court decisions concerning school prayer.

Objectives and Methods

This investigation seeks to develop a profile of those who favor and oppose prayer in public schools. We also wish to establish the relative importance of three variable clusters in predicting one's orientation toward this issue. Attitudinal, demographic, and religious variables are simultaneously used as predictors with discriminant function analysis. Broad religious groupings which reflect ideological differences are used as control categories to specify further the relative importance of the predictor variables within contrasting theological contexts.

SAMPLE

The data for this study are based upon three national samples. The National Opinion Research Center General Social Surveys for the years

---

[4] See the New York *Times*, May 7, 1982 at B40, Col. 1, and the March 1980 poll conducted by the Gallup Organization.

1974 and 1982 were chosen because they included the following specific question on school prayer:

> The Supreme Court ruled that is is unconstitutional to require the reading of the Lord's Prayer or Bible verses in public schools. Do you agree or disagree with this ruling?

The 1974 sample included 1484 cases and is based upon a modified probability sample with quota elements introduced at the final stage. The 1982 sample included 1506 respondents and its design is full probability, with predesignated respondents.[5]

The American National Election Study, 1980, serves as an additional data source. Below is the exact text of a school prayer question that was incorporated in the 1980 election survey:

> Which do you think—schools should be allowed to start each day with a prayer *or* religion does not belong in the schools.
>
> Should allow Depends Does not belong Other

The election study, conducted by the Center for Political Studies of the Institute for Social Research at the University of Michigan, represents eight discrete data collections, four of which involve a year-long panel and resulted in considerable missing data due to panel mortality. The sample size totals 3587. Tables 1 and 2, which profile the respondents in terms of school prayer, include data from all eight data collections; the discriminant analysis, however, is based on one of the last sets of data collected, thus eliminating much of the missing data associated with the long-term panel.

PREDICTOR VARIABLES

Background measures. The respondent's *age*, *years of educational attainment*, *sex*, and *race* were tapped as significant personal variables. Whether the individual ever has had *children* serves as a means to assess the possibility that attitude toward the issue is directly tied to its effect on the respondent's family. Present *region* of residence and *family income* were only utilized in the discriminant analysis. Three categories of region were entered: (1) Non-South, (2) southern border state, (3) solid South state.

Attitudinal measures. *Ideology* indicated the respondent's self-assessment on a seven-point liberal-conservative continuum. *Tolerance for the rights of atheists* (alpha = .77) was a composite measure based on three items which assessed the respondent's willingness to tolerate public speaking by atheists, to tolerate atheists teaching in a university, and to tolerate an atheist's anti-American book in the library.

---

[5] See the General Social Surveys, 1972–1982: Cumulative Codebook, James A. Davis, Principal Investigator.

Table 1. Prayer in Public Schools by Selected Attitudinal and Background Variables

	Proportion Favoring the Required Reading of the Lord's Prayer or Bible Verses		*Proportion Believing That Schools Should Be Allowed to Start Each Day with a Prayer*
	NORC General Social Survey, 1974 (N=1484)	*NORC General Social Survey, 1982 (N=1506)*	*American National Election Study, 1980 (N=3587)*
Total	67.9	59.7	72.1
Age			
Less than 30	55.0	54.2	66.1
31 to 55	73.1	63.6	75.5
56+	73.6	67.1	79.9
Education			
0 to 8 years	73.9	76.1	91.4
9 to 12 years	72.9	65.5	78.9
13+ years	51.1	49.7	59.8
Sex			
Female	66.2	58.0	75.9
Male	70.0	64.1	66.9
Race			
Nonwhite	67.3	59.6	83.4
White	75.3	75.4	71.4
Respondents who have children (any age)			
No	58.2	49.1	69.6
Yes	70.9	66.2	77.7
Political views			
Conservative	74.0	69.9	74.0
Moderate	70.0	60.9	70.4
Liberal	54.5	49.1	48.9
Tolerance for the rights of atheists			
Low tolerance	84.8	79.9	Not Asked
Moderate tolerance	71.4	68.5	
High tolerance	52.5	44.8	
Sex education in public schools			
Oppose	83.8	71.3	Not Asked
Favor	64.2	59.0	
Equal Rights Amendment			
Oppose	Not Asked	71.0	79.9
Favor		56.7	66.2

Table 2. Prayer in Public Schools by Selected Religious Variables

	Proportion Favoring the Required Reading of the Lord's Prayer or Bible Verses		*Proportion Believing That Schools Should Be Allowed to Start Each Day with a Prayer*
	NORC General Social Survey, 1974 (N=1484)	*NORC General Social Survey, 1982 (N=1506)*	*American National Election Study, 1980 (N=3587)*
Church or synagogue attendance			
Every week	77.8	70.7	86.8
Few times a month	74.2	67.4	79.9
Few times a year	59.3	56.4	69.9
Never	57.0	47.0	59.7
Extent to which religion provides guidance in life			
Great deal	Not asked	Not asked	85.1
Quite a bit			79.6
Some			71.3
Bible was			
God's work	Not asked	Not asked	86.4
Inspired by God			68.8
Man wrote			31.3
Religion			
Protestant	71.7	67.6	81.3
Catholic	73.5	58.8	75.5
Jew	22.2	23.5	26.2
Denomination			
Baptist	79.9	76.2	84.8[a]
Episcopalian	70.0	39.1	76.1
Lutheran	60.8	63.0	68.5
Methodist	67.0	64.2	77.9
Presbyterian	71.4	57.4	64.0
Sects	87.4	82.6	93.1

[a] Southern Baptist

Three other attitudinal variables were utilized from the NORC surveys. *Equal Rights Amendment* (not asked in 1974) support has been found to be related to political and religious liberalism (Tedin, 1978). Attitude toward *sex education* in the public schools was included because it too is a values-related public school issue.

Religion measures. The respondents' *church or synagogue attendance* was included. Data on measures of *religious guidance*, *Bible source*, and *importance of religion* were only available from the American National Election Study. Respondents were asked, "Would you say your religion provides some guidance in your day-to-day living, quite a bit of guidance, or a great deal of guidance?" The *Bible source* item asked the respondents to indicate if the Bible was God's work, inspired by God, written by man, or worth little.

Protestants were analyzed separately from Catholics with the multivariate model for the 1980 Election Study (Table 3). Additionally, respondents who reported a religious affiliation on the 1982 NORC survey were placed into one of four theological categories consistent with earlier studies (Stark and Glock, 1968, and Hadaway, 1978): (1) Liberal Protestants (Unitarian, Presbyterian, Episcopalian, United Church of Christ); (2) Moderate Protestants (Methodist, Lutheran, Christian Church, United Methodist, Wesleyan, Disciples of Christ); (3) Conservative Protestants (Southern Baptist, Baptist, Sects); and (4) Catholics (results not shown in table form). Jewish respondents were excluded from the multivariate analyses due to their small numbers.

Results

A majority of persons in the United States favor allowing prayer in the schools, but the percentage is dependent on how the question is worded. The American National Election Study included the milder of the ques-

Table 3. Multiple Discriminant Analysis for the 1980 American National Election Study

	All respondents (N=1333)			*Protestants (N=817)*			*Catholics (N=299)*	
	Total Cases	*Correctly Classified*		*Total Cases*	*Correctly Classified*		*Total Cases*	*Correctly Classifie*
Should allow schools to start day with prayer	840	644(76.7%)	1. Should allow prayer	567	76.9%	1. Should allow prayer	196	81.6%
Religion does not belong in the schools	493	316(64.1%)	2. Religion does not belong	250	63.2%	2. Religion does not belong	103	54.4%
% correctly classified (Hit ratio)		72.0%	% correctly classified (Hit ratio)		72.7%	% correctly classified (Hit ratio)		72.2%
Proportional chance criterion		53.4%	Proportional chance criterion		57.2%	Proportional chance criterion		55.1%
Standardized discriminant function coefficients			Standardized discriminant function coefficients			Standardized discriminant function coefficients		
1. Bible source (God = low)	+.34		1. Bible source (God = low)		+.41	1. Importance of religion (important = low)		+.33
2. Education	+.31		2. Age		−.32	2. Religious guidance (none = low)		−.29
3. ERA (approve = low)	−.26		2. Education		+.32	3. ERA (approve = low)		−.27
4. Importance of religion (important = low)	+.25		3. Region (Non-South, Border, South)		−.24	4. Ideology (liberal = low)		−.26
5. Ideology (liberal = low)	−.19		3. Religious guidance (none = low)		−.24	5. Bible source (God = low)		+.25
6. Religious guidance (none = low)	−.18		4. ERA (approve = low)		−.21	6. Age		−.23
7. Age	−.16		5. Importance of religion (important = low)		+.18	6. Education		+.23
7. Religious attendance (low = every week)	+.16							

tions in Table 1, only asking if the schools should be allowed to start each day with a prayer. Seventy-two percent agree that schools should be allowed to do so. The NORC question, on the other hand, is clearly within the Christian context, referring to the Lord's Prayer and Bible verses, and also includes the word "required." All children, Jewish as well as Christian and atheist, would be required to hear a Christian prayer. Not surprisingly, agreement with this question is somewhat lower than with the 1980 Election Study wording. The eight-year lag between the two NORC polls makes possible some inferences concerning trends in public opinion, although random variation could account for at least some of the differences. In the total sample, it can be seen in Table 1 that agreement has dropped from 67.9 percent in 1974 to 59.7 percent in 1982.

ATTITUDINAL AND BACKGROUND VARIABLES

Several trends are apparent in Table 1. First, as anticipated, older persons are more likely to approve of prayer in schools than are younger persons, especially those less than 30 years of age. Apparently, on this issue there is a clear division in the U.S. population between those under 30 and those over 30. We expected an age-related trend because of the tendency of older persons to be more politically conservative, but we may also be seeing the influence of another variable: having or having once had school-age children.

Persons with less education are also more likely to approve prayer in public schools. One can speculate that the liberalizing effects of education provide a broader context for assessing this issue. Looking next at sex and race, we see that the direction of the relationship differs from one poll to the other. In the NORC sample, males and whites are more likely to approve prayer in schools, while in the National Election Study, the reverse is true.

Whether a respondent has ever had children produces a relationship that is easier to interpret. Earlier surveys (see Princeton Religion Research Center, 1978) have shown parents, even many nonreligious ones, to be very concerned about religious training for their children. Many parents see religion as promoting "morality" and proper behavior on the part of children and they see it as a "good thing for kids" regardless of the religious content they receive.

A variety of political and social attitudes, also presented in Table 1, are associated with our dependent variable. There is a clear relationship between political ideology and favoring public school prayer. Persons who classify themselves as conservative or moderate in their political views are much more likely to approve of prayer in public school than are those who see themselves as political liberals. Further, it can be seen that the difference between liberals and moderates on this issue is much greater than between moderates and conservatives. Few people call themselves liberal, but those who do tend to be quite different from others on this issue. In a similar fashion, persons with more liberal views on rights of atheists, sex education, and the Equal Rights Amendment are more likely to disapprove of prayer in public schools.

RELIGIOSITY VARIABLES

It was expected that religious persons would be more likely to favor religion in the public schools than would nonreligious persons. And given the Christian orientation of the NORC item, we would also anticipate that Christians tend to favor such a change to a greater extent than Jews. Beyond the obvious, however, there are other significant questions.

The first three items in Table 2 focus on different aspects of religiosity: participation, salience, and orthodoxy. In all three, persons who are more religious are considerably more likely to favor returning prayer and Bible reading to the public schools. For religious attendance the relationship is quite strong and monotonic in both the NORC survey and the American National Election Study. As attendance increases, so does support for prayer in public schools.

Respondent's view of the Bible produces the strongest relationship of the three religiosity variables. A full 86.4 percent of those who believe the Bible is God's work wish to see prayer in the public schools; only 31.3 percent who see the Bible as the product of man favor it. Those who see it as God's message to man also believe that religion is needed and acceptable in many contexts of life other than the church and is especially needed in the schools.

Jews oppose prayer in the schools much more so than Catholics or Protestants. Indeed, they react just as negatively to the nonsectarian prayer in public-schools question on the 1980 Election Study as they do to the NORC item which refers specifically to a Christian prayer and the Bible. Perhaps Jews, given their minority status, assume that any prayer, scripture, or meditation is more than likely to be Christian or get translated in practice as Christian, and thereby to be offensive to them.

The analysis of denominational breakdown among Protestants is hindered by small subsamples, so closer attention has been given to the American National Election Study which includes a much larger sample than the NORC poll. Further, there is substantial agreement among Protestants that prayer should be returned to public schools. Variation within the list is present, however, and seems to be linked to religious conservatism. Those favoring prayer in public schools most strongly are also those who are most conservative: sects (93 percent in favor) and Southern Baptists (85 percent in favor). For the other four denominational clusters the conservatism connection is not as apparent. Methodists and Episcopalians do fall below Southern Baptists and are followed by Lutherans and Presbyterians.

Table 3 presents discriminant analysis results based on the 1980 Election Study for respondents who could be classified on the school prayer issue and on each of the variables utilized as discriminators.[6] This procedure generated a linear combination of standardized weighted variables which could be compared with respect to their relative importance in classifying a respondent on our dichotomous dependent variable into a category either favoring school prayer or opposing prayer in public schools.

First, an analysis of all usable respondents ($N=1333$) was conducted to assess the extent to which the hypothesized relationships were supported within a multivariate context. We had anticipated that religiosity items would be the most viable predictors because of their "most proximate" conceptual linkage with school prayer. A review of the lefthand portion of Table 3 reveals that the best single predictor is Bible source, a proxy variable for religious orthodoxy. The positive value(+.34) associated with this variable indicates that many of those respondents who believe that the Bible was God's work also contend that the school day should begin with a prayer. Three additional religion variables also appear in the list of key predictor variables.[7]

Advocates of school prayer, in addition to being theologically conservative and religiously active, tend to be less educated, to disapprove of the Equal Rights Amendment, are politically conservative, and are slightly older ($\bar{x}=45.6$ years versus $\bar{x}=39.4$ years) than those who believe that religion does not belong in the schools.

The standardized function was highly significant ($\chi^2=214$, df=13, $p<.001$) as were all the items comprising the function.[8] As can be seen, nearly 77 percent of those persons favoring school prayer were correctly classified while 64 percent of those in opposition were correctly placed. These overall classification results, particularly when considered relative to the proportional chance criterion percentage (53.4), are indicative of the variables' joint predictive power.

Results for the Protestant and Catholic respondents are shown in the middle and righthand portions of Table 3. While the "hit ratios" (72.7 percent and 72.2 percent) are nearly identical to the "hit ratio" for the combined analysis (72 percent), the rank order of the key predictor variables does differ in several respects. All variables entered as predictors, including several not shown which were not among the top seven variables, were statistically significant ($p < .001$). Again we see that school prayer advocates are far easier to predict, and for both the Protestants and Catholics the relative improvement over chance results (proportional chance criterion) is marked. Only among Catholic nonprayer advocates is the chance criterion larger than the percentage correctly classified (55.1 percent versus 54.4 percent).

---

[6] As noted earlier, Table 3 is based upon the eighth component (C3po) of the 1980 NES survey. A total of 1408 respondents were interviewed in this postelection stage of the survey. A document entitled "Design of the NES/CPS American National Election Study, 1980" provides details of the overall study design.

[7] The tolerance level was adjusted to minimize the possibility that correlated predictor variables would result in rounding errors in the discriminant coefficients and misclassifications.

[8] Several other items which are not shown in Table 3 because of their lack of importance as predictors were also included in the analysis. They were personal income, race, and sex. All were statistically significant.

Religiosity measures clearly dominate the list of key predictors in both instances; however, religious attendance has fallen from the list of centrally important predictors. We again see that the older, less educated respondents tend to favor school prayer, as do those who are socially and/or politically conservative.

We also placed the religiously affiliated NES into one of four theological categories as noted earlier. The results are not shown, as they tended to parallel those presented in Table 3. A similar analysis was also conducted on the 1982 NORC survey respondents (not shown).[9] Overall, the predictor variables in the NORC data set were less relevant to the school prayer issue than were those included in the 1980 NES survey. The hit ratios ranged from a low of 63.8 percent for Catholics to approximately 66 percent for the conservative, moderate, and liberal Protestant subsamples. We attribute the inability to predict accurately in these analyses to the relatively weak religion measures which were available.

Implications and Conclusions

Public policy does not always follow public opinion, even in a nation where elected representatives are supposed, ideally, to vote according to the wishes of their constituents. Polls have shown that the majority of the American public believes that schools should be allowed to start each day with a prayer, yet such activity has been banned by the federal courts for over 20 years.

Congress may continue to be asked to pass a prayer amendment. The fact that one has not been passed is undoubtedly related to many factors, not the least of which is the difficulty in writing an amendment acceptable to the disparate Christian groups who want the change. Other issues, however, are focused by the demographics and political orientation of those who favor or oppose the prayer amendments.

---

[9] The results which have not been shown are available from the authors.

Our data show that those who live in the South, have less education, and have lower incomes are more likely to favor prayer in the schools. It may well be that these social characteristics make this majority too silent to be heard by Congress, which itself is comprised of a highly elite group of individuals which our profile would classify as opponents to the issue. As we have seen, those with more than a high school education and of moderate to high incomes are *least* likely to favor returning prayer to public schools; members of Congress tend to be highly educated and come from affluent families.

Despite demographic obstacles, it seems that the ascendancy of conservative political thought in the United States would favor the passage of an amendment dealing with returning prayer to the schools. As can be seen in our profile in Tables 1 to 3, persons who favor returning prayer to the public schools tend to be consistently conservative over a wide range of political, social, and moral issues. Yet it would appear that this potentially large base of support may not be translated into legislation until representatives feel that prayer in the schools is something more than a lightly held belief among their constituents. Indeed, political scientists have long known that congressional perception of issue salience is a key intervening variable when considering the relationship between public opinion and legislative action (Miller and Stokes, 1963; Cnudde and McCrone, 1966).

Thus, the likelihood of a school prayer amendment probably depends on the ability of those favoring the change to mobilize their resources and present prayer in the schools as a potent political issue. Lobbying efforts are difficult to coordinate not only because of group differences in perspective but also because of congressional understanding that an "opinion gap" exists between Protestant leaders and lay persons (Hargrove, 1979:219).

References

Abington School District v. Schempp, 374 U.S. 203 (1963) in United States Supreme Court Reports: Lawyers Edition (1964). Rochester, N.Y.: Lawyer's Co-operative Publishing Co. Ser. 2, Vol. 10, 844–914.

Adams, W.
1975 "American public opinion in the 1960s on two church-state issues." Journal of Church and State 17:477–494.

Cnudde, C., and D. McCrone
1966 "The linkage between constituency attitudes and congressional voting behavior: a causal model." American Political Science Review 60:66–72.

Engel v. Vitale, 370 U.S. 421 (1962) in United States Supreme Court Reports: Lawyers' Edition (1964). Rochester, N.Y.: Lawyer's Co-operative Publishing Co. Ser. 2, Vol. 10, 601–619.

Hadaway, C.
1978 "Denominational switching and membership growth: in search of a relationship." Sociological Analysis 39:321–337.

Hargrove, B.
1979 The Sociology of Religion: Classical and Contemporary Approaches. Arlington Heights, IL: AHM Publishing Corp.

Lane, R.
1959 Political Life. New York: Free Press.

Miller, W., and D. Stokes
1963 "Constituency influence in Congress." American Political Science Review 57: 45–56.

Perkins, J., D. Fairchild, and M. Havens
1983 "The effects of evangelicalism on southern black and white political attitudes and voting behavior." Pp. 57–83 in T. Baker, R. Steed, and L. Moreland (eds.), Religion and Politics in the South. New York: Praeger.

The Princeton Religion Research Center and the Gallup Organization
1978 The Unchurched American. Princeton, New Jersey: The Princeton Religion Research Center.

Stark, R., and C. Glock
1970 American Piety: The Nature of Religious Commitment. Berkeley: University of California Press.

Stokes, A., and L. Pfeffer
1964 Church and State in the United States. New York: Harper and Row.

Tedin, K.
1978 "Religious preference and pro/anti activism on the Equal Rights Amendment issue." Pacific Sociological Review 21:55–66.

U.S. Constitution, Amendment I. (1791).

Way, H.
1968 "Survey research on judicial decisions: the prayer and Bible reading cases." The Western Political Quarterly 21:189–205.

Viewer Characteristics and Agenda Setting by Television News

DAVID B. HILL

DURING the recent past a fragile consensus has emerged among researchers which holds that television news programs can influence the relative salience of political issues in the minds of the American public—that is, that they are capable of setting the public agenda (Erbring, et al., 1980; Graber, 1980: 132ff; Iyengar, et al., 1982; MacKuen, 1981: 19–29; and McCombs, 1981). Generally, however, there have been few efforts to redress a problem noted by Iyengar (1979: 398), that "most previous agenda-setting studies . . . document the congruence between media content and issue salience without attempting to identify variables which strengthen or weaken the relationship . . . . " The research reported in this article responds directly to this criticism by examining the impact of viewer characteristics and modes of news exposure on the agenda-setting effects of television news.

Studies of television news recall and comprehension point to several factors that may influence the agenda-setting process. According to this

Abstract Most previous studies of agenda setting by television news have failed to consider factors that might strengthen or weaken agenda-setting effects. Drawing on the findings of news recall and comprehension research, this investigation of survey data gathered from 1204 national news viewers identifies several characteristics of viewers and viewing traits that increase the agenda-setting influence of news programs. The analysis suggests that effects are greater when viewers have prior awareness of news topics through print media exposure or have some college education. Watching news programs in color and attentively also contributes to increases in agenda setting.

David B. Hill, formerly Associate Professor of Political Science and Associate Director of the Public Policy Resources Laboratory at Texas A&M University, is a Vice-President of Tarrance and Associates, Houston. The author wishes to thank the John and Mary Markle Foundation and Drs. Ronald B. Frank and Marshall G. Greenberg for providing the data upon which this study is based. None of these parties bears any responsibility for the analyses or interpretations presented herein, however. This is an adapted version of a paper entitled "An Additive Effects Model of Exposure to Television News," which the author presented at the 1983 annual meeting of the Southern Political Science Association.

Public Opinion Quarterly Vol. 49:340–350 © by the Trustees of Columbia University
Published by Elsevier Science Publishing Co., Inc. 0033-362X/85/0049-340/$2.50

body of research, the recall and comprehension of television news is affected by (1) factors controlled by broadacsters, and (2) behavioral traits of the television news audience. Research into broadcaster-controlled factors has produced a plethora of contradictory findings regarding the efficacy of different methods of news presentation designed to foster viewer recall and learning. There seems to be agreement only that recall is enhanced whenever television viewers are more attentive to newscasts, but there is almost no consensus on how broadcasters can use various audio and video techniques to promote attentiveness (Stauffer, et al., 1983: 29–30; Woodall, et al., 1983).

Research into audience behavior and its bearing on news recall has produced significantly more promising results. There appears to be substantial evidence that the ability to recall news items is influenced by, among other factors: (1) viewers' motivations for viewing, whether for information gathering, diversion, or some combination of both; (2) viewers' prior levels of knowledge about the topic of a particular news item, or their interest in it; (3) viewers' attentiveness to the audio and video components of television news, especially when the two interact in some complementary fashion; and (4) the extent to which viewers plan in advance their exposure to news broadcasts (Graber, 1980: 134–44; Woodall, et al., 1983; Stauffer, et al., 1983; and Gantz, 1978).

Viewer characteristics and the manner of exposure to news programs should also influence agenda setting inasmuch as it is reasonable to assume that remembering and understanding the news precede the changes in attitudes that are inherent in conceptualizations of the agenda-setting process. If viewers' prior knowledge, motivations, and watching habits affect the recall of news items, then it seems plausible that these same factors should have some bearing on the process whereby viewers' issue agendas are established and modified.

Data

The data analyzed here were gathered as part of a larger study of the public's use of television. The methodology of that study is described in substantial detail by the co-principal investigators in their own report of their findings (Frank and Greenberg, 1980).[1] Study participants were interviewed between late 1977 and early 1978 and asked to complete an extended recall diary of all television viewing, on a program-by-program basis, for the fortnight prior to the interview. Thus the data consist of self-reports of the number of times, if any, that respondents watched national news programs during a two-week period. Furthermore, for each program watched during that period, respondents were asked to report on several traits of their viewing (e.g., attentiveness), again on a program-by-program basis. Finally, separate from the viewing diary portion of the study, respondents were asked more general questions about their attitudes toward television and other matters, including a series of items assessing interest in various news and public affairs issues.

---

[1] In addition to describing the sampling and interviewing procedures, this volume includes a complete copy of the various questionnaires used.

Data reported here are based on interviews with 1204 adult research subjects whose viewing diaries indicated that they had watched a national news program on at least one occasion during the two-week reporting period.

Hypotheses

We hypothesize that the agenda-setting power of television news will be influenced by several viewer characteristics and viewing traits. The first of these is the presence of a color television in the household. Because most news programs utilize color pictures and graphics to illustrate the news, and because some aspects of video are thought to enhance the recall of news items (Woodall, et al., 1983), we expect agenda setting to be stronger when news is viewed in color.

A second group of factors is related to how news watchers describe their own watching behavior. We predict greater agenda effects when: (1) viewers give the news their full attention without doing anything else at the same time; (2) viewers make a unilateral decision to watch the news (rather than watching after another member of the household chooses to watch the news); (3) viewers plan in advance to watch the news; and (4) viewers watch the news alone, not in the presence of others. Each of these expectations is consistent with notions that media effects are likely to be greatest when viewers actively seek out a program and watch it without distraction.

Other potentially important factors in news exposure are related to viewers' more general orientations toward television. We hypothesize that viewers' who perceive television as informative are more likely to have their agendas affected by television news exposure. When television viewing is associated with relaxation or escapist motivations, we predict reduced agenda-setting effects. While recreational viewing of television news may stimulate agenda setting on occasion, we would generally expect such exposure to be ineffectual. Collectively, these hypotheses reflect the expectation that agenda setting is apt to occur most strongly among viewers who perceive television to be a credible source of desired information and who use television to fulfill information-seeking goals. Conversely, we expect reduced agenda-setting effects when television viewing becomes a means of unwinding after work or is a vehicle for escaping from the cares of life.

Finally, prior research on learning from news exposure leads us to expect that prior awareness of news topics promotes agenda setting. While the present survey did not ascertain preexisting knowledge of *specific* news topics, it did measure formal educational attainment and exposure to print media news, including newspapers and the major weekly news magazines (*Newsweek*, *Time*, and *U.S. News and World Report*). Because exposure to college-level studies or to print news should have increased individuals' knowledge of the various topics discussed in news broadcasts, we hypothesize increased agenda-setting effects among respondents with some college education and those who reported reading the news sections of a daily newspaper or who subscribed to a major news magazine. Knowledge of news topics gained through education or reading is likely to promote a form of attentiveness and comprehension that enhances the agenda-setting influence of broadcasters.

Analysis of Agenda Setting

The methodology used here to measure and evaluate agenda setting is similar to that reported in other cross-sectional studies of the phenomenon.[2] The content of news programming prior to the interviews was content analyzed, and news topics were rank-ordered to establish the media agenda. This media agenda was then compared with the rank-ordered personal issue agendas of news watchers. Unlike many other studies that have measured personal agendas only in the aggregate, the survey data utilized here make it possible to compare media agendas with individual-level as well as aggregate personal agendas.

The measurement of the media's agenda was constrained to some extent by the individual issue salience questions. The latter were fixed choice items that asked respondents to rank (on a four-point scale from "extremely interested" to "not interested at all") their interest in 59 political and social issue topics. The Vanderbilt Abstracts of the three national network news programs (ABC, CBS, and NBC) were consulted to determine the number of news stories that made mention of each issue topic during the two weeks prior to the first interview day and throughout the remainder of the interviewing time frame. Only 20 of the 59 issue topics received substantial and consistent coverage during the time period investigated.[3] These 20 issues, shown in Table 1 with the number of news stories in which each topic was mentioned, were then rank-ordered from highest to lowest on the media agenda.

---

[2] McCombs (1981) has written the most sophisticated review of various methodologies used to explore agenda setting.

Aggregate personal agendas were determined by calculating the percentage of the sample that reported the highest level of interest ("extremely interested") in each social and political issue.[4] The issues were then rank-ordered according to those percentages to create the aggregate personal agenda. These rank orders are presented in Table 1. A comparison of the media agenda and personal agendas reveals only a weak relationship between the two. Of the 10 issues given most attention by the media, only three rank in the top half of the public's agenda.

A similar pattern of weak correlation was observed in the individual-level analysis. *Individual-level personal agendas* and their rankings were

Table 1. National News Program and Aggregate Personal Agendas

	Number of Stories	*Aggregate Personal Agenda Ranking*
Conflict in the Middle East	729	(15)
National economy	696	(7)
National unemployment	330	(8)
Sources and uses of energy	309	(5)
The stock market	235	(18)
Air transportation (planes, jets, etc.)	202	(17)
Agriculture and farming	140	(13)
Foreign policy	127	(19)
Labor unions in the U.S.	125	(12)
Medical sciences (anatomy, physiology, etc.)	124	(10)
Education and schools	113	(1)
Arms race	101	(14)
Space travel	101	(16)
Balance of trade	65	(20)
Managing money (finances, taxes, etc.)	56	(2)
Women's rights/roles	49	(9)
Problems of drug abuse	42	(4)
Health and nutrition	29	(6)
Causes and prevention of crime	20	(3)
Rights of minority groups	7	(11)

[3] The issue content of the national news programs was analyzed within every two-week period between September 15, 1977 and January 7, 1978. Issues that received heavy coverage during some blocks of time, but little or no coverage during other times, were excluded from the analysis. This technique has the advantage of tapping longer-term issue agendas of the media.

[4] Personal agendas are often measured by items that assess the perceived importance of various issues rather than the interest generated by these issues. While in many instances we would expect respondents to attach importance to interesting issues and vice versa, these two dimensions of issue salience (perceived importance and interest) are not likely to coincide for all issues. Because this study focuses on the interest dimension alone, the results reported may not generalize to the importance dimension of issue salience.

established using the four-point scale of interest in each social and political issue. The use of a four-point scale necessarily created numerous tied rankings of issue salience, a somewhat undesirable artifact of the fixed-choice question format as a means of assessing relative issue salience.[5] For example, if a respondent rated 2 of the 20 issues as ones in which he or she is "extremely interested," then those two issues would be assigned a tied ranking of 1.5 out of 20. Other ties were assigned numerical rank values by standard rank-ordering procedures (see Garson, 1976: 328–32). The media agenda for the 20 issues was then correlated with each individual-level personal agenda (and with the aggregate personal agenda where appropriate). The coefficients (values of Kendall's tau) resulting from these individual-level correlations will be referred to as the *agenda-setting score* for each respondent. The mean agenda-setting score for all news viewers was −.10, indicating a weakly negative relationship between the media and individual agendas. Thus, *among news viewers as a whole*, agenda setting is not confirmed by either the aggregate or individual analyses. It is possible, however, that agenda setting is more evident among news viewers with the characteristics and viewing habits described by the hypotheses stated previously.

First-order correlations between the individual level agenda-setting scores and viewer characteristics also produced values that are generally very small. The coefficients range from a null value (for watching alone) to .19 (for persons who are exposed to print news). The small magnitude of these correlations, along with the poor fit observed in the aggregate and individual level data, could be taken as evidence that agenda setting is not occurring, no matter what the nature of viewers' exposure to news programs. Such a conclusion would be unnecessarily pessimistic, however. In part, the small size of these correlations can probably be explained by research design constraints. A review of previous research suggests that cross-sectional studies such as this one reveal weaker effects than longitudinal studies. And by focusing on a set of 20 issues, rather than on a single issue (or two or three), the odds of achieving a mirror image relationship between media and personal agendas is significantly reduced. Furthermore, the measure of individual-level agendas is based on very elementary survey data which are likely to introduce a fair amount of "noise" into the analysis.[6] Therefore, we proceed with the analysis assuming that the actual extent of agenda setting is underestimated here. And because the *fundamental purpose of this research is to identify contingent conditions that affect agenda setting the most*, not to measure the precise dimensions of the phenomenon, the actual magnitude of agenda setting discovered here is relatively unimportant.

---

[5] Clearly, the many tied ranks of issue salience do not reflect respondents' real attitudes. For example, most respondents would be able to differentiate their levels of interest in each of the issue topics in which they were forced to report being "extremely interested." By assigning all such issues the same ranking, a substantial amount of measurement precision is lost.

[6] Each of these problems associated with cross-sectional survey approaches to agenda setting has been observed by others, including MacKuen (1981: 22), Erbring, et al. (1980: 17–21), and Bishop (1981: 609ff).

A comparison of the bivariate correlations between the agenda-setting scores and the independent variables indicates that the most important viewer characteristics are prior exposure to news topics, i.e., print media exposure (*r*=.19), some college education (.16), the presence of a color television in the household (.12), giving undivided attention to news programs (.12), choosing unilaterally to watch the news (.11), and planning in advance to watch the news (.10). Two other factors proved to be of slightly less importance: viewers' belief that television is informative (.07), and the fact that viewers frequently watched television with the intent of relaxing or escaping from everyday cares (−.07). The latter variable, as hypothesized, was negatively related to agenda setting. Watching news programs alone was unrelated (−.00) to agenda setting.

The relative importance of educational attainment and print media exposure suggests strongly that general news awareness influences the agenda-setting process. News awareness doubtless stimulates an attentiveness and a sensitivity to news items that does not occur among the less informed.[7] In short, the news-rich become richer while the news-poor lag behind. News awareness also facilitates superior comprehension and understanding of the news which, in turn, may affect the process of agenda setting.

While we hypothesized that viewing the news in color should enhance agenda setting, the relative strength of this variable, vis-a-vis the other independent variables, is surprising on account of the unpredictable effects of broadcaster-controlled factors in determining viewer recall of the news. Because a color presentation of the news involves some of the same variables manipulated by broadcasters (i.e., use of pictures versus "talking heads," spatial placement of image on the screen, use of film versus videotape, etc.) with inconsistent results, the impact of color television was not assured. This variable is also suspect in that the presence or absence of a color television in the household may simply reflect social status variations. This is particularly critical if the better educated are more likely to have color television, thus creating a spurious relationship between color viewing and agenda setting. The partial correlation between color viewing and individual agenda-setting scores, controlling for the effect of education, was .11, as compared with an uncontrolled coefficient of .14. Thus, the effect of color television is explained partially by the slightly greater incidence of color viewing by the college educated.

---

[7] The most often cited research to this effect is by Tichenor, et al. (1971). Their research indicates that the better informed about public affairs are apt to be more attentive to the media and as a consequence are more likely to be affected by it.

Because the bivariate analyses show that no single characteristic of newsviewing can unilaterally explain substantial variation in agenda setting, it is desirable to develop a multivariate model which combines the effects of several viewer characteristics in order to evaluate their collective impact on agenda setting. Multiple regression analysis was performed in order to determine the effect on agenda setting of various combinations of contingent conditions. A measure of the quantity of exposure (the number of times respondents watched a national news program during the prior two weeks) also was introduced into the analysis in order to determine whether quantity of exposure is as important a contingent condition as viewer characteristics and viewing traits.

Regression of viewing frequency and viewer trait variables on individual agenda-setting scores was conducted by a stepwise technique known as "maximum R^2 improvement." This method calculates all possible regressions of the independent variables on the dependent variable in order to find the strongest one-variable model, strongest two-variable model, and so forth, until a maximum solution is achieved.[8] Using this technique, we generated 10 different models of agenda setting. A "best" model was selected by taking the model with the fewest variables whose explanatory power (R^2) was not appreciably increased by the addition of additional independent variables. The six-variable model was judged best by these criteria. The addition of the remaining variables made almost no further contribution to explaining agenda setting.

Results of the regression analysis for the best model and the full 10-variable model are presented in Table 2. The small values of the coefficients doubtless reflect the same difficulties discussed earlier regarding the cross-sectional nature of the analysis and the level at which the indicators are measured. As we observed above, however, our major concern here is the comparative magnitude of the coefficients, not the absolute values. The results of the multivariate analysis add little substantive explanation to that provided by the simple bivariate correlations. The best model of agenda setting includes the four variables that were strongest in the bivariate analysis.

The most important result of the multivariate analysis has to do with quantity of news exposure. The regression exercise suggests that the frequency of exposure by news viewers is much less important than the characteristics of viewers and their viewing habits. All nine viewer characteristic variables outrank quantity of exposure as a predictor of

---

[8] The analysis was performed with the Statistical Analysis System (SAS).

Table 2. Results of Multiple Regression on Individual Agenda Setting Scores

	Best 6 Variable Model		*Full 10 Variable Model*	
	b	*Beta*	*b*	*Beta*
Intercept	−.14		−.18	
Print media exposure	.07*	.15	.06*	.14
Full attention given to news programs	.06*	.14	.04*	.11
College education	.04*	.10	.03*	.09
Color TV in household	.02*	.08	.02*	.07
TV judged informative	.01*	.08	.01*	.08
Relaxation/escape motivation for viewing TV	−.01*	−.07	−.01*	−.07
R chose news programs			.03*	.08
Plans viewing of news programs			.02	.08
Watches news programs alone			−.02	−.04
Frequency of news viewing			.01	.03
	$R^2 = .08$		$R^2 = .08$	

* *F* significant at .01.

agenda setting. In order to evaluate this finding further, the six "best model" viewer characteristic variables and frequency of exposure were regressed together on the absolute level of interest in each of the 20 issues ranked highest on the agenda of television news, i.e., 20 separate regression analyses with 7 independent variables in each. *In every single instance, frequency of exposure to national news programs failed to be a significant positive predictor of interest in the news topic.* In 10 of the resulting equations, frequency of exposure was negatively related to interest in the topic. These were not significant, however. Results of 2 of the 20 regression analyses are presented in Table 3 for illustrative pur-

Table 3. Results of Regression on Interest in the Middle East and the National Economy

	Middle East—Beta	*National Economy—Beta*
Print media exposure	.18*	.18*
College education	.16*	.13*
Full attention given to news programs	.10*	.09*
Frequency of news viewing	.04	.00
Color TV in household	.03	.01
Relaxation/escape motivation for viewing TV	−.02	−.05
TV judged informative	−.01	−.05

* *F* significant at .05.

poses. The standardized regression coefficients presented in the table indicate that print media exposure, college education, and paying full attention to news programs are the strongest and only significant predictors of interest in these news topics. While frequency of viewing is not the weakest predictor in both equations (it is, in fact, the fourth strongest predictor of interest in the Middle East), it is not a significant variable in any of these equations. Across the 20 equations, frequency of exposure was the weakest or next to the weakest independent variable in a majority of cases. Thus, after news viewers achieve a minimum threshold of quantitative exposure, their personal characteristics and viewing habits, particularly their news knowledge and attentiveness to news programs, are more important in establishing agendas than mere time spent in front of the TV set.

Discussion

This study has substantiated some of the conclusions and speculations of previous studies of agenda setting. For example, the rather weak indications of agenda setting reported herein are consistent with the notion that it is more difficult to demonstrate the phenomenon in cross-sectional surveys, particularly when considering attitudes toward as many different issues as were considered here. Longitudinal experiments and quasi-experiments designed to probe the impact of television news on the salience of one particular issue or class of issues (e.g., foreign policy matters) probably reveal more robust effects.

This study has also confirmed the apparent importance of "news awareness" (indirectly indicated by print media exposure and educational attainment) in conditioning the effects of exposure to television news. While previously only MacKuen (1981) had formally scrutinized the effect of educational attainment on agenda setting, the role of news awareness as a contingent condition of news recall and comprehension has been fairly well established. In this context, perhaps the most significant aspect of this study has been to argue that education and other variables which affect news recall and comprehension may similarly affect agenda setting. In order for the two literatures to be properly reconciled, however, future research should directly ascertain the effects of news recall and comprehension on the process of agenda setting.

Another important finding of this study concerns the *relative* impotency of quantity of television news exposure as a factor in agenda setting. This result has implications not only for the study of agenda setting, but also for the study of other media effects. As studies by Levy (1978) and Hill (1983) seem to indicate, quantity and quality of news viewing are two independent concepts. And whereas prior research into almost all media effects has focused on the quantity of exposure alone, it may be that researchers need to focus now on some of the relevent characteristics of viewers and their viewing traits other than frequency of exposure.

References

Bishop G.
1918 "Survey research." In D. Nimmo and K. Sanders (eds.), Handbook of Political Communication. Beverly Hills: Sage Publications.
Erbring L., E. Goldenberg, and A. Miller
1980 "Front-page news and real world cues: a new look at agenda-setting by the media." American Journal of Political Science 24: 16–49.
Gantz, W.
1978 "How uses and gratification affect recall of television news." Journalism Quarterly 55:664–681.
Garson, G.
1976 Political Science Methods. Boston: Holbrook Press.
Graber, D.
1980 Mass Media and American Politics. Washington, D.C.: Congressional Quarterly Press.
Frank, R., and M. Greenberg.
1980 The Public's Use of Television. Beverly Hills: Sage Publications.
Hill, D.
1983 "Qualitative dimensions of exposure to political television." Social Science Quarterly 64:614–623.
Iyengar, S., M. Peters, and D. Kinder
1982 "Experimental demonstrations of the 'not so minimal' consequences of television news programs." American Political Science Review 16:848–57.
Iyengar, S.
1979 "Television news and issue salience." American Politics Quarterly 7: 395–416
Levy, M.
1978 "The audience experience with television news." Journalism Monographs 55.
MacKuen, M.
1981 "Social communication and the mass policy agenda." In M. MacKuen and S. Coombs (eds.), More Than News. Beverly Hills: Sage Publications.
McCombs, M.
1981 "The agenda setting approach." In D. Nimmo and K. Sanders (eds.), Handbook of Political Communication. Beverly Hills: Sage Publications.
Stauffer, J., R. Frost, and W. Rybolt
1983 "The attention factor in recalling network television news." Journal of Communications 33:29–37.
Tichenor, P., G. Donohue, and C. Olien
1971 "Mass media flow and differential growth of knowledge." Public Opinion Quarterly 34:159–70.
Woodall, W., D. Davis, and H. Sahin
1983 "From the boob tube to the black box: television news comprehension from an information processing perspective." Journal of Broadcasting 27:1–22.

Finding Subgroups for Surveys

ROGER TOURANGEAU AND A. WADE SMITH

In 1982, the General Social Survey (GSS) participated in a methodological experiment to compare two samples of black respondents. One sample consisted of black respondents identified through a household sample designed as a cross-section of the general population; the other consisted of black respondents identified through a household sample in which areas were selected with selection probabilities proportional to the size of the black population. The study was funded by the National Science Foundation and conducted by NORC. Respondents in both samples were administered a face-to-face interview that lasted about 75 minutes

Abstract This article reports on an experiment comparing two methods for selecting national samples of blacks. The experiment was carried out as part of the 1982 General Social Survey (GSS). One sample was selected using area probability sampling with the selection probability for an area proportional to the size of its black population. The other sample was selected by sampling extra dwelling units in areas selected for the regular GSS. Although comparable in other respects, the two methods differed sharply in costs and statistical efficiency. The sample selected from the special black frame proved less expensive to interview and required fewer screening interviews in order to locate households with black members. The sample selected in the same areas as the regular GSS sample exhibited a smaller "design effect"—probably because fewer respondents were selected from each sample area. The decision to use one method or the other to sample rare subgroups would depend on several factors, including the degree to which the subgroup is segregated, the amount of savings possible through piggybacking, and the relationship between screening and interviewing costs.

Roger Tourangeau is Senior Methodologist, National Opinion Research Center. A. Wade Smith is Associate Professor, Arizona State University, and Research Associate, National Opinion Research Center. This research was supported by a grant from the National Science Foundation (SES-8023396) to the National Opinion Research Center. The authors especially thank Mary Cay Burich for providing the accounting data used in these analyses, and Martin Frankel, June Meitz, and two *POQ* reviewers for their insightful comments on earlier drafts. The authors take responsibility for all conclusions herein, as well as any remaining errors. Direct reprint requests to: Library, National Opinion Research Center, 6030 S. Ellis Avenue, Chicago, IL 60637.

Public Opinion Quarterly Vol. 49:351–365 © by the Trustees of Columbia University
Published by Elsevier Science Publishing Co., Inc. 0033-362X/85/0049-351/$2.50

and included the regular GSS questionnaire with additional questions on issues pertaining to blacks.

From the standpoint of sampling methodology, the design of a national sample of blacks is an instance of the more general problem of sampling "rare" populations for which there is no list, or sampling frame. The usual sampling frames for national household surveys are lists of dwelling units on preselected sample blocks; for telephone surveys, the usual frame is a list of area codes and exchanges to which randomly generated suffixes can be appended. In either case, only a minority of the selected units will contain respondents who are eligible for a survey of blacks. Eligibility is generally determined by a short preliminary interview, or "screener." Since blacks constitute roughly 10 percent of the general population, if the sampling frame has been developed as a cross-section of the general population, then an average of 10 or more screening interviews will be necessary to identify a single eligible household.

The need to screen large initial samples can have a dramatic impact on the cost of a survey. For this reason, sampling methodologists have developed a number of strategies for reducing screening costs. The GSS experiment compared two of these strategies. NORC developed a special sampling frame, in which blocks with larger numbers of blacks had higher selection probabilities. In addition, in the same areas where interviews were being conducted for the regular GSS, additional dwellings were selected and screened for the Black GSS. The screening was conducted in conjunction with the data collection for the regular GSS in an attempt to reduce the costs of the screening. The two samples will be referred to as the *black frame* and the *black oversample*.

All of the strategies developed by sampling statisticians share one of two basic goals. One is to increase the "take" rate of the screening operation. The take rate refers to the percentage of screeners which yield members of the eligible population. The other is to reduce the costs of conducting the screening interviews. After a general discussion of the major techniques that have been used to select samples of rare populations, we describe and compare the design and selection of the two GSS samples of black respondents.

Methods of Sampling Rare Populations

Kish (1965:404–415) distinguishes at least eight separate techniques for sampling rare populations; it will serve here to describe five basic methods with a few of the more important variations on each one. The basic methods include piggybacks or amalgams, large clusters, batch testing, disproportionate sampling, and networking. These categories are not mutually exclusive and can be combined in particular sample designs.

All five of these special techniques should be contrasted to the most common sample design, in which a large cross-sectional sample is selected and screened.

Piggybacking refers to the use of data collected for another survey for screening purposes. For example, if a large general population survey is already planned, it is often quite inexpensive to add a few screening questions to locate the members of a rare population eligible for a second survey. Piggybacking is contemplated more often than it is used, owing to the practical difficulties it entails. Successful piggybacking requires that data from one survey be available to the researchers conducting another, and confidentiality requirements often prevent such exchanges. Piggybacking also generally requires close coordination of the data collection schedules for the two surveys—it is not very helpful to know that a black family resided at a particular address three to five years ago. A technique that is closely related to piggybacking is the amalgam approach, in which several rare characteristics are covered in the same screening interview; the same screener may locate blacks for one survey, elderly for another, and immigrants for a third.

The aim of the piggyback and amalgam is to reduce screening costs on a per case basis. They share this goal with a second sampling technique—the selection and screening of large clusters. In most survey designs, the size of the clusters is determined as a compromise between statistical efficiency and cost per case. (Cluster size refers to the number of interviews conducted in each of the next-to-the-last stage sampling units; in area probability sampling, these units are usually referred to as *segments*, a convention followed here.) Larger cluster sizes tend to reduce the cost per case of data collection; on the other hand, because elements within a cluster tend to be similar to each other, larger clusters reduce the precision of sample results. Across a range of survey topics and survey organizations, the compromise is often struck at about five cases per cluster, the number used in the regular GSS and in many of the recurring surveys conducted by the Census Bureau (such as the Current Population Survey; see Bureau of the Census, 1978). Cluster sizes for screening samples are often several times larger than this, since relatively few cases from each cluster will be eligible for the main sample. For example, if the expected eligibility rate is only one in 10, then a cluster size of 50 for the screening operation will yield an average cluster size of five for the main interview. The drawback to this procedure is that it can produce sharp variations in the final cluster sizes: some clusters may contain no members of the eligible population, while others may consist entirely of eligible members of the target group. These variations in the cluster size increase the difficulty of staffing the data collection effort and also decrease the statistical efficiency of the sample.

Batch testing is a method that can reduce the cost per screener even further. Batch testing takes its name from quality control operations in which batches rather than individual articles are tested. The analogous procedure in survey research applications is to screen a single unit in each cluster in order to determine whether to retain the cluster for further screening. For example, a single dwelling unit is screened on a sample segment. If that unit contains a member of the population of interest, other units on the segment are also screened; if the unit does not include at least one eligible member, no further screening is done on the segment. This procedure, originally developed for use in random-digit-dialing samples (Waksberg, 1978), has been extended to area probability sampling (Sudman; undated); it is most efficient when the members of the eligible population are highly clustered and when a substantial proportion of segments include no members of the eligible population.

Disproportionate sampling and networking are designed to increase the take rate, or the number of eligible units located per screener. In disproportionate sampling, strata are systematically overrepresented in the sample if they contain higher concentrations of the target populations. For example, in an area sample, areas with high percentages of blacks would be placed in a single stratum; the sampling rate would be higher in this stratum than in strata consisting of areas with fewer blacks. Kish (1965:409–410) has shown that this technique is of surprisingly limited usefulness, unless the bulk of the eligible population is concentrated in a few strata in which the take rates are substantially higher than those in the remaining strata. In addition, oversampling requires the use of data weights that compensate for the unequal selection probabilities; the use of weights typically reduces the statistical efficiency of sample estimates. Sampling with probabilities proportional to the size of the target population, the technique used to select the black frame, is closely related to the strategy of oversampling. In both techniques, the selection probability of a unit is related to the size of the eligible population—but sampling with probabilities proportional to size (PPS) has the advantage that it usually produces an equal probability, or self-weighting, sample (see, for example, equations [1]–[4] below).

When PPS methods are used to select an equal probability sample for screening, the expected take rate is necessarily the same as the expected take rate for any other equal probability sample, including samples selected as cross-sections of the general population. By definition, every equal probability sample gives the same chance of selection to each unit in the population; within the limits of sampling error, no one group is overrepresented relative to any other group. Sampling with probabilities proportional to the size of the eligible population does have two advantages over other equal probability samples. First, it reduces the variability in the cluster sizes—in this case, the number of blacks in each segment. A sample with roughly equal cluster sizes is easier to staff and yields more efficient estimates than a sample with clusters that vary sharply in size. Second, small departures from equal probability sampling can produce large efficiencies in screening. Even within the context of PPS sampling, some areas will be selected in which the eligible population constitutes a small proportion of the total population. By reducing the sampling rate within such areas, the overall take rate can be improved without a sharp reduction in statistical efficiency.

Networking and related techniques attempt to increase the take rate by increasing the number of units screened in each screening interview. It is typical in screening operations for a single informant to provide screening information on all persons who reside at the same dwelling unit. In networking (Sirken, 1970), the respondent in the screening interview provides information not only about the members of his or her own household but also about the members of linked households; the respondents may, for example, be asked questions concerning the eligibility of persons in the households of their parents or siblings. Although it is convenient to associate each element in the population with a single sampling unit (for example, to associate each person in the target population with a single dwelling unit), it is not essential. Networking schemes link individuals to multiple sampling units—a person may be linked with his or her own household and the households of first-degree relatives—and for this reason the technique is often referred to as *multiplicity sampling*.

The difficulty with networking techniques is that for many characteristics the screening data will be less accurate than the data from a more conventional screening sample. Errors in the screening data can produce serious sampling bias, since eligible members of the target population will be excluded from the final sample. On the other hand, networking can improve the take rate, often dramatically, allowing for a larger sample of members of the target group. A more conventional technique that is related to networking is to allow a single respondent to give data on all the members of a compact cluster; respondents, for example, may be asked screening questions concerning their next-door neighbors as well as their own households. Clearly, this variation on networking is only useful when the characteristics that determine eligibility are highly visible; otherwise, response errors in the screening may produce unacceptable bias in the final sample.

Much of this discussion is summarized in Table 1 below, which indicates the usual strengths and weaknesses of each procedure.In terms of this classification of techniques, the black frame is an example of PPS sampling with probabilities tied to the size of the *eligible* population and

Table 1. Techniques for Sampling Rare Populations

Procedure	*Strengths*	*Weaknesses*
Piggybacking (related method: amalgam)	Reduced cost/screener	Coordination of field schedules; confidentiality
Large clusters	Reduced cost/screener	Variation in final cluster sizes reduces efficiency
Batch testing	Increased take rate	Population must be highly clustered
Disproportionate sampling (related method: PPS)	Increased take rate	Unequal weights reduce efficiency (not a problem with PPS)
Networking	Increased take rate	Response error in screeners

so is related to the general method of disproportionate sampling. The black oversample, on the other hand, used a method related to piggybacking, in which the screening operation was linked to the data collection for the regular GSS.

Method

THE BLACK FRAME

The special frame developed for the black GSS consisted of a sample of 51 blocks where NORC staff compiled a complete enumeration of the dwelling units. The blocks were selected using a three-stage stratified sampling procedure. The selection of individual dwelling units on the sample blocks constituted the final stage of the selection of the screening sample.

The units for first stage selections (PSUs) consisted of standard metropolitan statistical areas (SMSAs) or, in less urbanzied parts of the country, individual counties. The entire set of PSUs was stratified according to geography and level of urbanization and then was further ordered according to the black population as a percentage of the total PSU population. A systematic sample of 17 PSUs was then drawn with the selection probability for each PSU proportional to the size of its black population. Letting Pr_{1i} refer to the selection probability for PSU i and MOS_i to its measure of size, the selection probability becomes

$$Pr_{1i} = (17 \times MOS_i)/MOS \qquad [1]$$

MOS in equation [1] represents the total measure of size across all the PSUs—that is, the black population of the United States. (Seventeen represents the number of PSUs selected.)

Three secondary units were selected within each of the sample PSUs. In rural areas, the secondary units consisted of Census Bureau enumeration districts; in more urbanized areas, they consisted of block groups.

These second-stage selections were made using systematic sampling with selection probabilities proportional to size; the black population in the secondary unit served as the measure of size. If Pr_{2ij} denotes the selection probability for secondary unit j in PSU i and MOS_{ij}, the measure of size for that unit, then the second-stage probability is

$$Pr_{2ij} = (3 \times MOS_{ij})/MOS_i \qquad [2]$$

Three is the number of secondary selections per sample PSU. Fifty-one secondary units were selected in the 17 sample PSUs.

The units for third-stage selection are called segments and consists of blocks in urban areas and block-like units in rural areas. One segment was selected from each of the 51 second-stage units in the sample. As with the first- and second-stage selections, the third-stage units were sampled systematically with selection probabilities proportional to the black population. The selection probability for segment k in second-stage unit j in PSU i (Pr_{3ijk}) was therefore

$$Pr_{3ijk} = MOS_{ijk}/MOS_{ij} \qquad [3]$$

MOS_{ijk} refers to the segment measure of size and MOS_{ij} to the measure of size for the entire second-stage unit.

Within a sample segment, the selection probabilities are the same for all dwelling units. For the dwelling units in segment k in second-stage unit j in PSU i, the selection probabilities depend on the measure of size for the segment and desired number of selections per segment, b^*, which was set at about 11:

$$Pr_{4ijk} = b^*/MOS_{ijk} \qquad [4]$$

Although b^* is a constant that applied to all 51 sample segments,[1] the actual number of selections varied considerably across segments, since the measure of size for the segment was not the same as the number of dwelling units. (The expected number of selections in a segment is the within-segment sampling rate, as given in equation [4], times the number of dwelling units.) In all, 1400 dwelling units were selected from the black frame.

The overall selection probability for a particular dwelling unit is the product of the selection probabilities at each stage (described by equations [1]–[4]). It is evident from the equations that the overall probability reduces to a constant—every dwelling unit had the same chance of being selected into the sample. In one segment, however, the

---

[1] Pr_{4ijk} cannot exceed 1.0; for this reason, every segment had to meet a minimum measure of size requirement. Segments with eligible populations smaller than the minimum were combined with adjoining segments until the combined unit exceeded the minimum size requirement.

expected number of black households was quite low relative to the total number of dwelling units; within this segment, the sampling rate (Pr_4) was reduced by half. Thus, for the dwelling units selected in 50 of the segments the overall selection probabilities were identical and can be denoted by *k*; for the dwelling units in the remaining segment, the selection probabilities were equal to one-half *k*.

THE BLACK OVERSAMPLE

The second sample of blacks was selected from the frame for the regular GSS. This sample consisted of additional dwelling units in the same segments where the regular GSS interviews were being carried out. The sample design for the regular GSS is, in many ways, parallel to the design used to select the black frame. The regular GSS sample is a multistage area probability sample in which counties and SMSAs are selected in the first stage, block groups and enumeration districts are selected in the second stage, segments are selected in the third stage, and individual dwelling units are selected in the final stage. Two separate selections of dwelling units were made: the first consisted of 2222 dwelling units designated for the regular GSS interview; the second consisted of 1500 dwelling units designated for a short screening interview to determine whether any blacks resided there. Whenever a dwelling unit is selected for the regular GSS, it is excluded from further sampling. For this reason, there is no overlap between the regular GSS sample and the black oversample.

SUMMARY

Table 2 summarizes the procedures used to select the special black frame and the black oversample and highlights the differences between them. There are three main differences between the two designs: additional stratification was used in selecting the segments shared by the regular GSS and the black oversample; the sample sizes at each stage of selection of the oversample are larger than those for the special black frame; and the measure of size used in selecting the black oversample was based on total rather than black population. The most important of these is the difference in the measure of size. Because it uses black population as the measure of size, the black frame consists of a sample of dwelling units on a set of blocks specifically selected to represent the black population; the black oversample, by contrast, is a sample on a set of blocks chosen as a cross-section of the entire population.

Table 2. Sample Design Summary

	Black Frame	*Black Oversample/ Regular GSS Frame*
First stage		
Number of selections	17	101
Unit	County, SMSA	County, SMSA
Measure of size	Black population	Total population
Stratification	Census division, urbanization	Census division, urbanization, size, % black
Second stage		
Number of selections	51 (3 per PSU)	303 (3 per PSU)
Unit	ED, BG	ED, BG
Measure of size	Black population	Total dwelling units Family income, % black
Third stage		
Number of selections	51 (1 per ED, BG)	303 (1 per ED, BG)
Unit	Segment (block)	Segment (block)
Measure of size	Black population	Total dwelling units
Fourth stage		
Number of selections	1400	1500 black oversample (2222 regular GSS)
Unit	Dwelling unit	Dwelling unit

Results

There are a number of criteria on which sample designs can be evaluated: take rates, relative costs, sampling bias, nonresponse, sampling errors, and response errors are some of the most important. Data are available to compare the black frame and the black oversample on the take rates, relative costs, nonresponse, and sampling errors. Although there is little basis for a comparison of sampling bias or response errors, there is also little reason to expect that the two samples differ on these criteria. This expectation receives support from the results of the analyses presented by Smith (1984), which demonstrate that estimates based on the two samples do not differ significantly. Differences in the amount or direction of sample bias or response error should be reflected by differences in the results from the two samples.

The take rates for the black frame and the black oversample differ sharply. Leaving out unoccupied dwelling units, the take rate is 25.5 percent for the black frame (344 eligible households out of 1349 occupied dwelling units) versus 10.7 percent for the black oversample in the regular GSS frame (150 out of 1407). In part, the difference between the take rates reflects the adjusted sampling rate in one of the black frame segments. It can be seen from equation [4] that in segments with a small black population the within-segment sampling rate (i.e., Pr_4) can approach 1.0, with the result that nearly every dwelling unit in the segment would be included in the screening sample. This situation arose in one segment in the black frame, where nearly 650 dwelling units were initially selected for screening; the initial sampling rate in this segment (as defined by equation [4]) was subsequently reduced by one half. The weighted take rate for the black frame, which adjusts for this departure from equal probability sampling, is about 20 percent.

The relative field costs associated with the two methods also differ sharply. It is difficult to separate screening costs from the costs of the main interview; the figures presented here combine both types of field cost. For the black frame, each completed interview required, on the average, about eight hours of interviewer labor and $18.00 of expenses (mainly travel costs). The black oversample required 16 hours of interviewer labor and over $28.00 in expenses. There are several reasons for these differences, some of them grounded in the design itself. Although similar numbers of screeners were completed in the two samples, more than twice as many interviews were completed in the black frame. The dwelling units selected for screening in the black frame were highly clustered with an average of 27 dwelling units in each segment as opposed to an average of five in the black oversample. The dense clustering made for more efficient use of interviewer labor during the screening operation. Because some of the questions for black respondents raised sensitive issues, only black interviewers were used to conduct the main interviews in both samples. In the black frame, it was reasonable to use only black interviewers for both the screeners and the main interviews. In the oversample, selected in the same segments as the regular GSS sample, the majority of the interviewing staff (and the majority of the respondents) were white. When a black respondent was located in the oversample, a black interviewer was assigned to complete the main interview, increasing the number of callbacks and the amount of travel time. Thus, the efficiencies that might have been gained from piggybacking the black oversample and the regular GSS data collection were, to a large extent, lost.

The cost figures reported here should be interpreted cautiously, since they do *not* include frame development and other start-up costs. Frame development, in this context, refers to the process of selecting and listing area segments. These costs can be substantial, although exact figures from the GSS are not available. It is important to realize that both samples incur such costs. The black oversample was selected as a subsample of NORC's "master sample," a large area sample that had been selected and listed at the beginning of the decade. The black frame was selected from the listings compiled in areas selected specifically for this study; in principle, these listings could also have been used to select additional samples of blacks since the sample selected for the GSS experiment included only a small portion of the listed dwellings. Because several samples can be drawn from the same set of listings, it is not always clear how to allocate frame development costs among the surveys or how to allocate them on a per case basis.

The two samples show very similar response rates. In both samples, about 97 percent of the screeners were completed; interviews were completed with about 71 percent of the households at which eligible respondents were found.

Because the black frame included a larger number of cases than the black oversample, estimates based on the black frame will be more precise (that is, they will have smaller standard errors) than estimates based on the black oversample. In order to eliminate the difference in sample sizes, the two samples will be compared using a measure of relative efficiency, the design effect. The design effect compares the precision of an estimate from a complex sample design (such as the stratified, multi-stage designs used here) with the precision of an estimate based on a simple random sample of the same size as the actual sample. It is the ratio of the actual variance of the estimate to its variance under the assumption of simple random sampling. A larger design effect indicates a less efficient sample. The design effect is sensitive to a number of design features, including unequal weights and average cluster size.

For 11 statistics from each sample, design effects were calculated using a computer program, UVAR, developed at NORC. The statistics were proportions based on demographic variables, such as sex and marital status. Data weights were computed to compensate for differences in individual selection probabilities and both weighted and unweighted data were used in the comparison. Because the weighted and unweighted results show only minor differences, only the weighted results will be presented here. Across the 11 statistics, the average design effect is 2.1 for the black frame and 1.6 for the black oversample. This is a substantial difference in efficiency. On the average, the black frame is as precise as a simple random sample consisting of 117 completed cases; the oversample has the same precision as a simple random sample of 68 cases. The difference in statistical efficiency is a byproduct of the denser clustering of the black frame respondents. These 247 respondents are concentrated in 17 PSUs and 51 segments. By contrast, the 107 respondents from the black oversample are spread throughout more than 40 PSUs and more than 100 segments. The impact of clustering is even more dramatic for variables, such as geographic region or size of place, which do not vary within a PSU. The design effects for estimates involving such PSU-level characteristics consistently exceed 10.0 for the black frame but range from 2.2 to 4.2 for the black oversample. In calculating average design effects for the two samples, PSU-level variables were excluded.

Despite the difference in statistical efficiency, the black frame retains the advantage in terms of cost efficiency. If the field costs are re-expressed in terms of cost per effective case—the black frame producing 117 effective cases for estimates based on individual-level characteristics and the black oversample producing 68—the black frame is still more cost-efficient. An average of 17 hours of interviewer labor were needed for each effective case in the black frame versus 25 hours in the black oversample. Table 3 summarizes the comparison of the two samples.

Table 3. Comparisons Between the Samples

Characteristic	*Black Frame*	*Black Oversample*
Sample sizes		
Initial sample	1400	1500
Occupied dwellings	1349	1407
Completed screeners	1312	1366
Eligible households	344	150
Completed interviews	247	107
Take rate	25.5%	10.7%
Completion rates		
Screener	97.3%	97.1%
Main interview	71.8%	71.3%
Cost per completed case		
Interviewer hours	8	16
Other expenses	$18	$28
Statistical efficiency		
Design effect	2.1	1.6
Mean cluster size	4.8	<1
Effective sample size	117	68

Discussion

Because both samples used the same questionnaire and field procedures, it is no surprise that they differ most sharply in their take rates and costs. What is surprising is that the black frame was superior in both areas—it produced a higher take rate and a lower cost per case. It seems clear, however, that the differences in costs must be attributed, in part, to the decision to match interviewer and respondent race in the black oversample. The matching made it difficult to share field costs with the regular GSS, which was being conducted at the same time and in the same segments. It is not unusual to attempt to match the demographic characteristics of interviewers and respondents; a number of studies have suggested that matching can produce improvements in data quality (see Bradburn, 1983, for a summary). Thus, the cost results for the black oversample are probably more representative of the method of screening a sample selected from a cross-sectional frame than of the method of piggybacking. In situations where interviewer time (and costs) can be shared by two surveys, the costs of screening a cross-sectional sample might be considerably reduced.

The Black GSS demonstrates the utility of the PPS approach for sampling blacks but suggests that a number of factors may tilt the balance in favor of oversampling or some other approach. What general guidelines can researchers use in deciding between the methods? A closer examination of the cost structure of a survey will help make the answer clearer.

Equation [5] gives a simple model of the costs for a two-phase survey (in which n_s households are first screened and n_I eligible cases are identified and interviewed):

$$C = n_s C_s + n_i C_I \qquad [5]$$

where C refers to the total cost of the survey, C_s to the average cost per screening interview, and C_I to the average additional cost of completing the main interview. If one case is selected for interview from each household with an eligible member, the screening sample (n_s) must be $1/t$ (where t is the take rate) times larger than the final sample (n_i). For this reason, we can express the total cost per completed case as

$$C/n_I = C_s/t + C_I \qquad [6]$$

When the total cost of the survey is fixed, the best design will be the design that maximizes the final sample size by achieving the lowest cost per case. Equation [6] makes several points readily apparent. First, it shows that reducing the cost per screener (C_s) has exactly the same impact as increasing the take rate (t); for example, halving the screening cost has the same effect as doubling the take rate. Equation [6] also makes another point clear. When the marginal costs of an interview (C_I) are large, then changes in the screening cost (C_s) or take rate (t) will have relatively little effect on the total costs per case. The black oversample results confirm this point. Matching of interviewer and respondent race apparently increased the marginal costs of an interview so much that any savings in screening costs were lost.

The framework of equation [6] suggests that the decision to use PPS sampling or piggybacking should rest on three factors: (1) reductions in screening costs through piggybacking, (2) increases in the take rate through PPS sampling, and (3) the impact of either method on the cost of the final interview. The GSS experiment underscores the importance of the last two factors.

The black frame sample—selected via PPS sampling—had a take rate more than twice as high as that of the black oversample. This gain generally depends on adjustments to the within-segment sampling rate defined in equation [4]. As Waksberg (1978) has shown in another context, PPS sampling without adjustment produces the same take rates expected from a simple random sample. Adjustments will prove effective whenever the sample can be expected to include some segments with very low expected take rates. Rare groups that are segregated geographically (e.g., racial and ethnic minorities, the rich, the poor) are thus good candidates for PPS sampling. Because PPS sampling requires an estimate of the size of the eligible population in each segment, researchers can estimate beforehand whether the population is sufficiently segregated that adjustments to PPS sampling are likely to produce gains.

The GSS experiment also demonstrates the importance of the marginal costs of the main interview. While the black oversample may have decreased the cost per screening interview, it apparently increased the marginal costs of the main interview. The black frame allowed the same interviewers to conduct both screening and final interviews—sometimes in a single trip to the field—reducing the marginal costs. When the main interview requires a separate interviewing staff (because of special training needs or matching requirements), gains from piggybacking the screening costs are likely to be offset by increases in the costs of the main interview.

Precision is not strictly a matter of sample size. Both PPS sampling and piggybacking generally produce clustered samples with design effects greater than 1.0. For this reason, equation [6] should be modified in practice to reflect costs per effective case:

$$DC/n_i = C_s/t + C_l \qquad [7]$$

in which *D* refers to the design effect. A number of features of the sample design influence the design effect; for clustered designs, the most important is generally the average cluster size. Formulas are available for estimating the impact of clustering (Kish, 1965:162) and for finding the optimal cluster size (Waksberg, 1978). In assessing the relative merits of piggybacking and PPS designs, researchers should take into account the expected design effects under both designs.

References

Bradburn, N.
1983 "Response effects." In P. Rossi and J. Wright (eds.), The Handbook of Survey Research. New York: Academic Press.
Bureau of the Census
1978 The Current Population Survey: Design and Methodology. Technical Paper 40.
Kish, L.
1965 Survey Sampling. New York: John Wiley.

Sirken, M.
1970 "Household surveys with multiplicity." Journal of the American Statistical Association 71:257–266.
Smith, A. W.
1984 "Evaluating the products of alternative sampling methods." Paper presented at the 1984 American Sociological Association Meetings, San Antonio, Texas.
Sudman, S.
Efficient Screening Methods for the Sampling of Special Populations. University of Illinois, Survey Research Laboratory. Undated.
Waksberg, J.
1978 "Sampling methods for random digit dialing." Journal of the American Statistical Association 73:40–46.

Matching Survey Responses to Official Records: An Exploration of Validity in Victimization Reporting

PETER V. MILLER AND ROBERT M. GROVES

THE RECORD CHECK study has a prominent place in the literature on survey methodology. In seeking to validate self-reports of "matters of fact" and to test the efficacy of different data collection techniques, researchers in a variety of fields have sought to ascertain the degree of match between survey reports and external record evidence for events of interest. The results of these investigations have shaped not only data collection procedures, but also widely held beliefs about the mechanisms underlying error in self-reports. In this article, we scrutinize the procedures of the record check study itself in an analysis of results from recent research on victimization reporting. Our findings offer a different focus for interpretation of record check results. We call for a more complete, standardized procedure for reporting record check findings and for a more critical view of their results.

Abstract Record check studies—involving the comparison of survey responses with external record evidence—are a familiar tool in survey methodology. The findings of a recently conducted reverse record check study are reported here. The analyses examine match rates between survey reports and police records, employing more or less restrictive match criteria—e.g., using various computer algorithms versus human judgments. The analyses reveal marked differences in the level of survey–record correspondence. Since the level of match rate appears highly variable depending on the definition of a "match," we advocate reexamination of the "lessons" of previous record check studies which employed only vaguely specified match criteria. We argue, further, that record evidence may best be employed in constructing alternative indicators of phenomena to be measured, rather than as the arbiter of survey response quality.

Peter V. Miller is Associate Professor of Communication Studies and Research Faculty, Center for Urban Affairs and Policy Research, Northwestern University. Robert M. Groves is Associate Research Scientist, Survey Research Center, Institute for Social Research, University of Michigan. Research for this article was partially suppprted by a contract from the Bureau of Justice Statistics, U.S. Department of Justice. The article does not necessarily represent the views of the Department of Justice. The authors are indebted to Allen H. Andrews, Superintendent of Police, Peoria, Illinois, for making the study possible and to Dr. Charles Cannell, Dr. Charles Cowan, and Dr. Wesley Skogan for insightful comments on an earlier draft. The authors are responsible for any errors which remain.

Public Opinion Quarterly Vol. 49:366–380 © by the Trustees of Columbia University
Published by Elsevier Science Publishing Co., Inc. 0033-362X/85/0049-366/$2.50

Background

The use of record evidence to check self-reports has formed the basis for a number of discussions of survey error. For example, Parry and Crossley (1950), in an often cited article, used records of voter registration, voting, library card holding, and contributions to a charitable fund to check reports of those activities from a survey. Similarly, Bradburn et al. (1979) attempted to match survey reports of voter registration, voting, library card holding, bankruptcy, and drunk driving to records of those activities. The findings of these studies suggested that socially desirable activities (e.g., voting) tend to be overreported—more reports of such events were found in surveys than could be located in records—and that undesirable activities or states (bankruptcy) tended to be underreported. Bradburn et al. interpret the findings of their matching analysis as the effects of "question threat" on self-reports.

"Administrative" record check research in other fields has led both to alterations in data collection techniques and to a further buttressing of the conclusions from the studies already cited. Methodological research which preceded the inception of the National Crime Survey (NCS), for example, included three record check studies using police records (Yost and Dodge, 1970; Dodge, 1970; and Turner, 1972) which examined various survey procedures for gathering reports of victimization. During the course of these investigations, the NCS questionnaire was modified substantially, and beliefs were formed about the amount and type of error in victimization reporting. Like the previously reviewed research, these studies noted that crimes tended to be underreported when compared with police record evidence and that certain events, such as assaults, were more likely to be underreported than were less serious episodes, such as larcenies.

Study Questions

Police records, like hospital, doctor, or voting records have had the surface appeal of a tangible criterion against which to compare survey reports.[1] A contributing factor to this impression has been the unproblematic way in which the act of matching a survey report and a record of the event has been described in the record check literature. In the "typical" investigation, matching is done by human comparison of the survey protocol and the record. Once matching is complete, matched and unmatched cases are compared analytically on characteristics such as date, type of event, etc. The results of such analyses are the basis for claims such as, "nonsalient events are underreported." These analyses rest on the ability of judges unambiguously to assign "match" or "non-match" designations to every survey case. The match exercise, however, is not without complexity: the coder must determine whether the person interviewed is the same person named by the record and then test the correspondence between event details expressed in each source. In both tasks, there are a variety of more or less restrictive criteria which can be employed to assess the degree of fit. Thus, the matching step represents a set of decisions about relevant criteria for comparison and allowable values on each criterion.

---

[1] In this article, we set aside serious questions which may be raised concerning the standard treatment of record evidence as "truth" in record check studies. A number of writers—e.g., Marquis (1978), Skogan (1976, 1981)—have critiqued the assumption that record evidence is free of error and that mismatches between survey reports and records are always proof of error in the self-reports.

It is rather surprising, therefore, to see virtually no space in reports of record check studies devoted to this crucial step. The nature of a match is left undefined in most reports. While some degree of intercoder agreement is sometimes required, the procedures which led to consensus are not specified. No mention is made of the study rules which may subject coders to implicit pressures to find matches or to disagree at the slightest provocation. No comparisons are made of match rates derived from human decisions and ones derived from other means, such as computer algorithms. Finally, matches are considered "dichotomous" propositions: one match rate using one set of (unspecified) criteria is presented, rather than a *range* of match rates based on various definitions of correspondence. Producing a single match rate based on a set of ill-specified human decisions gives the illusion that match decisions are unproblematic observations and shifts the focus of inquiry to a consideration of the flaws in survey reports.[2]

In the analyses to follow, we present the findings of a reverse record check of victimization reporting as a range of match ratios derived from machine and human matching processes, and using more or less restrictive, reasonable match criteria. We illustrate the highly variable nature of match ratios depending on the considerations which go into constructing them. Our analysis compares the ratios produced via various machine algorithms with the one produced through human judgments. Our findings lead us to question the nature of match ratios reported in previous record check studies.

---

[2] We do not mean to imply that the matching procedures in all previous record check studies have been haphazard. Rather, matching procedures in such studies have not been a focus of inquiry, and consequently there has been little attention paid to reporting details of the procedures. In victimization research, see, for example, Turner (1972), Sparks, et al. (1977) and Schneider (1977) for a comparison of the varying levels of specificity with which matching procedures are reported.

Study Design

Approximately 1300 police records for January–September 1981 were sampled from the Peoria, Illinois, Police Department files. During October and November, 1981, interviews were sought from all persons at least 12 years old in the households attached to the telephone numbers listed for the victims of the crimes. In addition, some 570 telephone numbers in the Peoria area were generated by RDD (Random Digit Dial) methods and called to obtain interviews. These cases were added to the sample in order to conceal from the interviewers the fact that known crime victims were being called.[3] In the following analysis, we consider only the police record households.

Approximately 370 of the sample police records had nonworking or nonresidential telephone numbers; other records reflected multiple victimizations of the same person or of several persons simultaneously. When all these facts were accounted for, about 900 distinct working household numbers were linked to sample police records.

An interview was completed with at least one person in approximately 765 of these households, resulting in a household level response rate of 85 percent. A total of 1577 persons were interviewed in the police record households, and we estimate that this represents about 75 percent of all persons at least 12 years old in those households.

PROCEDURES FOR MATCHING RESPONDENTS WITH POLICE RECORD VICTIMS

The first type of matching problem we want to consider is the overlap between the people listed on the police records and the people we interviewed by calling the phone numbers listed for them on the records. Our intent in the matching process was to reflect levels of uncertainty in determining whether the victim listed by the police matched an interviewed person. To this end, on each variable listed as a match criterion, levels of certainty of the match were recorded. This permits the categorization of people as definite matches, definite nonmatches, or as one of several intermediate categories. Matching people in records and survey reports is the most basic sort of match undertaken in record check studies; it may seem straightforward at first blush, but it can be complex indeed, as seen in the following match rules.

Using the information collected during the interview, three variables were chosen for use in person-level matching: (a) first name, (b) age, and

---

[3] This procedure attempts to reduce the likelihood that interviewers might obtain more victimization reports merely because they know that each sample case reported a crime. Not all record check studies have used this design. In some, namely Dodge (1970), the interviewers knew that all sample cases were listed as victims on police records. In other studies, e.g., Sparks, et al. (1977), extensive "salting" with nonrecord cases was done.

(c) gender. The age match used year of birth as reported in the police record and the interview record. The gender match used the recorded data on the police record and the interviewer observation in the interview case.

The name match produced the largest difficulties. A sampling of various kinds of cases indicates the problems faced by the matching process:

Misspellings (e.g., "Wilbur" vs. "Wilber")
Possible second names (e.g., interview "Harold," record "Z. H. Smith")
Possible initials (e.g., interview "Harold," record "H. Z. Smith")
Name not listed on the interview (e.g., interview "Husband," record "Thomas Ziegler")
Possible different names for same person (e.g., interview "Maude," record "Mrs. John Terpes")
Qualifications to names (e.g., interview "Jerome, Jr." vs. record "Jerome Johnson").

To deal with these classes of person identification comparisons, we implemented the following four-category code for name matches:

1) Definite match: includes exact first name matches, and misspellings judged to be misspellings, e.g., "Dorathy" and "Dorothy"; nicknames, e.g., "Mike" and "Michael"; qualifications to names, e.g., "Frank, Jr." and "Frank, Sr." Both are coded as "1."
2) Potential matches: includes initials, e.g., "D. J." vs "Donald"; no first names, e.g., "Mr. Thompson" vs. "Norman Thompson"; possible second names, e.g., "Paul" vs. "Irvin P. Barr."
3) No name listed in interview, e.g., "Sister," "Household informant," "No Name."
4) Definite nonmatches: includes cases with names on both records that are clearly different, e.g., "Donna" vs. "Robert."[4]

Three variables, age, gender, and name, were combined to determine a match status group that consisted of 13 different categories. For each police-listed victim, assignment to one of the 13 match categories was performed with reference to the respondents associated with the victim's telephone number. As shown in Table 1 categories 1−12 are devoted to those cases for which listed victims and at least one respondent at the listed phone number are of the same sex *and* have the same reported age within nine years *and* are not definite name mismatches. Others are classified into group 13; thus, group 13 consists of all cases where police record victims and respondents do not match on gender *or* are definite name mismatches *or* have reported age differences of 10 or more years.

---

[4] We must note that person matching in this study was somewhat more difficult than is probably the case in other record check investigations because of the amount of name information we had to work with. Because of the need to conceal from the interviewers the fact that we were using a victim sample, we followed standard SRC procedures for household listing, which do not permit collecting information on respondents' last names. Previous record check studies which had full name information undoubtedly did not encounter the same level of complexity in matching persons. On the other hand, some of these investigations have been criticized for using name/address samples of victims, possibly cuing the interviewers that the listed person should have something to report.

In households with multiple respondents, there were only six cases in which more than one household member could have been coded as a "possible match" (codes 1−12) with the victim. In all other such households, all but one household member were easily assigned to code 13.

Table 1. Match Categories for Police Report Victims with Interviewed Households

	Match Category			
Group Number	*Name Match*	*Age Difference*	*Number of Victims*[a]	*% of Victims*
1	Definite match	0−1	361	47.0%
2	Potential match	0−1	13	1.7
3	Probable nonmatch	0−1	91	11.9
4	Definite match	2	2	.3
5	Potential match	2	0	0.0
6	Probable nonmatch	2	0	0.0
7	Definite match	3	2	0.3
8	Potential match	3	0	0.0
9	Probable nonmatch	3	2	0.3
10	Definite match	4−9	2	0.3
11	Potential match	4−9	0	0.0
12	Probable nonmatch	4−9	6	0.8
13	Definite nonmatch or different gender	10 or more or different gender	287	37.4
Total			766	100.0

[a] There were 765 different telephone numbers sampled from police records, where at least one person provided an interview. One of these households contained two sample victims, and therefore there are 766 victims who could have been a respondent.

Table 1 shows that the most restrictive definition of victim match would result in 47 percent of the police-reported victims from interviewed households having a match with a respondent in that household (definite name match, gender match, age within one year). If we ignore name as a criterion and emphasize sex and age, over 60 percent of the reported victims match with a police record.

Comparing these match rates with previous victimization reverse record checks, we note that our most conservative estimate of person match is roughly the same as the "interview completion rate" (43 percent) reported by Sparks et al. (1977). Our most liberal estimate is slightly less than the "completion rates" of the three NCS pretest record checks, which ranged from 63 percent to 69 percent. All of these previous studies began with *name* samples from police records, and interviewers were instructed to locate the named persons. We gave interviewers a *telephone number* sample and directed them to take interviews with all eligible respondents (12+ years of age) at eligible (noncommercial working) numbers. The person-matching process in our study, therefore, was more difficult than that of previous investigations. The point here, however, is that in addition to sampling frame, the nature and stringency of the match criteria can markedly affect person match estimates. Additionally, it is interesting to note that, across studies, attempts to locate victims listed on police records result in finding about one-half to two-thirds of them (assuming that "matched" cases constitute found victims).[5] The relatively low level of success in finding victims listed on police records provides another sobering context for record check study interpretation.

Incident Level Matching Procedures

Another way to compare our work to past record check studies is to examine the variability in *incident* match ratios across a wide range of match criteria. That is, to what extent do police reports of particular events match survey reports of incidents which matched respondents say they have suffered? How is the match level affected by various match rules? Considering only definitely matched persons (Group 1 of Table 1), two general types of *incident* match procedures were used; the first were completely computer-based rules based on recorded characteristics of the incidents; the second were human judgments based on the written summaries of incidents in police records and interview schedules.

COMPUTER-BASED MATCHING RULES

Three variables recorded both on the police record and on the interview incident form were used for the computer matching procedures: type of crime code assigned to the incident, date of the incident, and location of the incident. Type of crime codes were assigned to interview incidents using the same procedures used in National Crime Survey data processing. The type of crime code used in the Peoria police record system was a four-digit code which was generally compatible with the coding used in NCS. Exceptions to this compatibility existed in the record coding system's lack of distinction between crimes with and without injury to the victim. A collapsing of the two coding systems to create a single compatible code of 23 categories was performed (see Miller et al. 1982).

---

[5] Even in the case of the Chicago Police Department's (1983) investigation of its own records of "unfounded" crime, 40 percent of the investigated records yielded no determination as to whether the crime report had been appropriately listed as baseless or not, since the "victims" in these cases could not be located. The problem of finding victims is intimately related to the very definition of crime, as the Chicago investigation reveals, since without a located victim no crime can be said to have occurred.

For matching on the date of the incident, the month of occurrence of the incident was asked of each respondent who reported a victimization. Similarly a date was recorded by the responding officer in the police records.

Matching on location of the incident is somewhat more complicated. The NCS incident report form (which we employed in this study) employs a field-coded question to determine the location of the crime. The answer categories are:

1. In own home
2. In detached structure near home
3. Vacation home
4. Near home
5. Public transportation
6. Friend's, neighbor's home
7. On street, in motor vehicle
8. Restaurant, bar
9. Office, factory
10. Other commercial building
11. Apartment parking lot
12. Other parking lot
13. School building
14. School property
15. Park, playground
16. Other
17. Don't know

Police record information on location of the crime was not coded numerically in the record, unlike type of crime and date information. SRC coders were assigned the task of coding the location of the crime described by police verbally in the records using the code employed in the interview schedule. Although the records were generally fully detailed on the crime location, there were some difficulties involving the categories "in own home," 'near own home," and "on the street." Where no reasonable solution could be found, missing data values were assigned.

With coding complete, an automated count was performed of the incidents that had the same values on the three variables—type of crime, date, and location. To measure the sensitivity of the results to variation in complexity of the coding, different aggregations of the codes in the three variables were introduced. For the type of crime coding, for example, some match ratios are based on a 23-category code; others on a 4-category code. Similarly, a 17-, 4-, and 2-category code were compared for location of the incident. For the date variable, contrasts are made between reports in the same month, within one month of one another, and within three months of one another.[6]

What results from the various collapsing rules are 24 different criteria for determining if there is a match between the incident report and the police report. The most stringent match rule requires the two reports to have the same month of occurrence, the same location on the full 17-category coding, and the same crime type on the 23-category code. The figures presented in Table 2 are match ratios: the number of incidents mentioned in the interview with the same values on the match variables as the police record. If more than one interview incident matches in a single person's interview, the interview would contribute two units to the numerator. Thus, it is possible that the ratio can exceed 1.0.

Using the most stringent rule the match ratio is .136; only a few of the crimes sampled are matched by reports by persons interviewed in the survey. That ratio can be almost doubled by relaxing the rule to include those interview reports dated a month before and a month after the police record date (analysis number 2, with a .236 match ratio). In summary, analyses 1–19 in Table 2 use different numbers of categories on the location and type of crime variables and different dating requirements. Using the three variables in all of their forms produces a range in match ratios from .136 to .736. The latter match ratio uses a two-category location code ("in or near own home" vs. "somewhere else"), a four-category type of crime code ("physical assaults," "larceny," "burglary," and "other"), and the most liberal date code (three months before or after the police report). The use of the three match variables alone thus can produce a wide range of match ratios; full understanding of a match ratio requires knowledge of the details of the coding of the match variables.

Past record check studies may not have used all three variables simultaneously; indeed, the most frequent strategy appears to be one in which reports are matched on incident characteristics exclusive of the date of the victimization. (The "matched" reports then are analyzed for dating accuracy). Analyses 20 through 24 present match data using the "no date" rule. If, for example, no date criterion is used but the most stringent location and type of crime code criteria are used, the match ratio is .36. If, however, the two-category location and four-category crime type coding are used (the loosest match coding), we find a .90 match ratio.

When the machine match criteria are further limited to only a single variable, type of crime, the most disaggregated crime type coding (with

---

[6] While we clearly advocate full reporting of match procedures, editorial considerations rule out a complete description of these codes. See Miller, et al. (1982) for details.

23 categories) yields a .77 ratio of matched crimes to police reports. The most liberal match criterion, using only a four-category type of crime match, yields a ratio over 1.0. The ratio greater than 1 results from the fact that some victims reported more than one incident with the same type of crime code as the police reported crime.

To summarize the computer-based matching analyses, we have seen that match rates can be constructed using only three criteria, with different values on each, which range from about .14 to a nearly perfect fit. These efforts to create empirical rules for matches between interview reports and police records yield less than satisfactory results to a researcher seeking to evaluate the nature of reporting error. This dissatisfaction stems from the ease with which the match ratio can be manipulated by changes in explicit criteria for matching. With the lack of documentation on matching procedures that is common in the literature on record check studies it is difficult to interpret the findings of past studies. Inference from these studies is a function of the reported match ratios; match ratios are a function of matching criteria. Without analysis of the sensitivity of the reported ratios to the criteria, any inference is subject to challenge.

USING HUMAN JUDGMENT FOR MATCH DECISIONS

A problem with computer match procedures concerns the limited set of variables on which matches can be made. Incident reports, both in the police record and in the interview, have theoretically thousands of different characteristics. Few of these are coded on both data sets in such a way that comparisons can be made. At the end of any such exercise with computer matching, questions still remain unanswered about the nature of the correspondence between the interview report and the police report. Two reports can match on two, four, or ten characteristics and still be reports about *different* incidents. Machine matching procedures in some cases appear poor substitutes for human review that can simultaneously consider many variables *and* utilize any other information for matching that may be available on only a subset of the police records or interview reports. This selective supplementation of match criteria so easily performed by human review has no doubt led many past researchers to use human judgment to produce match decisions.

To provide some comparison to past studies, a "manual match" procedure was performed in this study. The procedure had research assistants compare the written summaries of the incident. On the police record there was a written summary that was entered by the reporting officer and was sometimes updated after the initial report with new information about the incident. At the end of the incident report in the interview, the

Table 2. Incident Match Results for Various Incident Match Criteria Among Matched Victims[a]

	Incident Match Criteria			*Victim Interview Reports Matching the Sampled Police Report*	
		Number of Categories in Match Variable			
Analysis No.	*Difference Between Months*	*Location of Incident*	*Type of Crime*	*No.*	*Ratio Matching Interview Reports/ Police Reports*
1	0	17	23	33	.136
2	1	17	23	57	.236
3	0	4	23	50	.207
4	0	17	11	43	.178
5	1	4	23	86	.355
6	1	17	11	71	.293
7	0	4	11	60	.248
8	1	4	11	102	.421
9	3	17	23	71	.293
10	3	4	23	113	.467
11	3	17	11	87	.360
12	3	4	11	135	.558
13	0	17	4	49	.202
14	1	17	4	79	.326
15	0	4	4	73	.302
16	1	4	4	123	.508
17	3	17	4	96	.397
18	3	4	4	167	.690
19	3	2	4	178	.736
20		17	23	87	.360
21		4	4	207	.855
22		2	4	218	.901
23			23	186	.768
24			4	278	1.149
25	0	manual	match	87	.360
26	≤1	manual	match	134	.554
27	≤3	manual	match	144	.595
28	match on	manual	code	173	.715

[a] The case base for Table 2 ($N=242$) is definitely matched victims (Group 1 from Table 1), *excluding* those cases where the information on sampled police records did not permit adequate comparison of survey reported incidents and police recorded incidents. The excluded cases are ones, for example, where the person recorded on the police record was found and interviewed, but the police record contained too little information to judge whether or not the incident in question matched the survey report. Also excluded were cases where the incidents on the police record were commercial crimes since such incidents were not covered by the survey questionnaire. Thus, Table 2 only includes definitely matched victims whose survey reports were "matchable" to police recorded incidents. Incidents are weighted to remove the few duplicate reports of the same incident by a respondent.

interviewer typed in a summary of the events in the incident. These summaries often described the relationship between victim and offender, gave a short sketch of the events of the crime, and described effects for the victim (e.g., what property was stolen, what injuries were suffered). The location of the crime was often listed, but no dates of the events were recorded. In the vast majority of cases these summaries were the only information used by the research assistants to judge whether the two reports described the same event. In a small number of cases where there were two or more interview incident reports that were very similar, the dates of the police record incident and the interview incident were compared to inform the decision. Also in such situations and in cases where the police record did not list all of the salient details of an incident, supplementary police files were checked to see whether a match existed. For example, if an interview recorded that certain items were stolen, but the police record did not name the items specifically, the "property record" files were reviewed for the victim in question to distinguish between similar incidents or to determine if a single interview incident matched the police report.

The rule used in determining whether two reports described the same event could be termed one of omission versus exclusion. Two descriptions were judged to describe different events if there were some contradictions (e.g., a tape deck stolen from a parked vehicle without any damage to the car versus a tape deck stolen by first breaking out the side window). Two descriptions were judged to describe the same event if there appeared to be omissions in one of the reports that would not lead to a logical contradiction (e.g., a tape deck *and* some cassettes stolen from a car by first breaking a window versus a tape deck stolen from a car by first breaking a window). The judges worked together discussing each case. They were instructed to be absolutely sure about a match decision when it was made. Problem cases were turned over to a supervisor for resolution.

Table 2 presents match ratios based on the manual match procedures. The match procedure itself (without concern for a date match criterion) produces a match ratio of .715 (Analysis 28). This is quite comparable to the match ratio achieved by the 23-category type of crime code alone (with no constraint on date or location). The ratio is almost double that achieved by the combination of a 17-category location and a 23-category type of crime code.

If consideration of the reported date of the incident is added to the analysis of the manual match (Analysis 25), then the match ratio declines, to .36 for reports in the same month. It increases to .554 for those within one month, and to .595 to those within three months (Analyses 26 and

27). One can note that this pattern of increase in match ratio by relaxing the date criterion is similar to the findings of Sparks and others (1977: Table III.4) that when all "telescoped" incidents and accurately dated ones are combined, the match rate is some 40 points higher than if only accurately dated incidents are counted in the numerator.

Discussion and Conclusion

The foregoing analyses have illustrated the malleability of "matches" in record check studies. We have seen that, depending on the criteria employed, the values allowed on them, and the procedures for matching (machine vs. human), we can produce match ratios which range from .14 to around 1.00. These findings lead us to question the common practice of measuring and reporting a single match rate in record check studies. For the sake of replicability alone, it is clear that such investigations should produce a carefully defined *range* of match statistics. Such a practice would more fully inform other investigators about the ways in which self-reports depart from record evidence.

Further, our match approach raises some conceptual issues about the nature of matches in record checks. For example, is a misdated event merely "telescoped," or is it possibly not the same event as the one in the record file? Standard procedure in previous studies has been to match incidents irrespective of date, and then to analyze the amount of "telescoping" that occurred in reporting. While in many cases respondents only have one event of a given type to report within a reference period, some crime victims suffer multiple, similar incidents. For these people, code named "series victims" in the National Crime Survey, we cannot be sure that a "telescoped" report is not, in actuality, a report of a *different* crime of the same type. This problem is particularly acute in victimization record checks, since it is believed that series victims account for a large portion of all victimization events, but it may also arise in other substantive areas (e.g., doctor visit record checks).

It may be argued that the variability in match results is not so great in areas such as health, voting, or financial reporting, and that we have emphasized the disparity in match rates in our selection of criteria and code collapsing rules. We suspect that victimization record checks may evidence higher levels of match variability than, say, hospitalization studies. However, this is an empirical question and we hope that data can be brought to bear on it. As for the design of our machine match analyses, which do show widely different match ratios, it appears that the most stringent match criteria we employ are more restrictive than those used in past studies, while the most liberal standard is probably more liberal than ordinarily would be employed in a match analysis. We must emphasize, however, that it is not clear what criteria have been used in most past studies. Further, it is interesting to place the "human" match results in the context of the machine match findings and to note its closer fit to the more liberal rules we devised.

While record check studies typically do not specify match criteria carefully, this is not to say that the problem of matching itself has escaped the attention of survey statisticians. Considerable work has been done by government statisticians in Canada and the United States in response to the need for the merging of large data files involving, for example, Census and Social Security information. The matching research has focused on issues ranging from theoretical models for matching (e.g., Fellegi and Sunter, 1969) to pragmatic discussions of matching assumptions and procedures in federal studies (e.g., Office of Statistical Policy and Standards, 1980). This work is directly relevant to the problem of matching in record check studies since it addresses the nature of error in matching and desiderata in selecting match characteristics.

But the problematic nature of match analysis has been ignored in previous record check studies in favor of a focus on the "bottom line"—the extent to which survey data depart from "truth," as measured by the record evidence. Reporting the variability in match rates—by expressing uncertainty about the fit between the two kinds of evidence—reinforces the view that record evidence does not represent unassailable truth. Records of events may have more value in providing alternative indicators of the object of survey report, rather than as means for validating the reports. Record samples, in addition, offer substantial practical advantages for methodological research on rare events such as crime. Thus, while we do not believe that record check studies offer a panacea for assessing error in survey data, we feel that such designs can provide important insights about the product of different data collection schemes. Our call for a different focus on match results is not an attack on the method itself.

References

Bradburn, N., S. Sudman, and Associates
1979 Improving Interview Method and Questionnaire Design. San Francisco: Jossey Bass.
Chicago Police Department, Bureau of Administrative Services, Auditing and Internal Control Division
1983 "Detective division reporting practices." Internal report.
Dodge, R.
1970 "Victim recall pretest—Washington, D.C." Washington: U.S. Bureau of the Census, memorandum.

Fellegi, I., and A. Sunter
1969 "A theory for record linkage." Journal of the American Statistical Association 64:1183–1210.
Marquis, K.
1978 "Record check validity of survey responses: a reassessment of bias in reports of hospitalizations." The Rand Corporation.
Miller, P., R. Groves, and V. Handlin
1982 "Peoria reverse record check study: initial data analysis." Research report, U.S. Bureau of Justice Statistics.
Office of Federal Statistical Policy and Standards
1980 "Report on exact and statistical matching techniques." Statistical Policy Working Paper 5. U.S. Department of Commerce.
Parry H., and H. Crossley
1950 "Validity of responses to survey questions." Public Opinion Quarterly 14:61–80.
Schneider, A.
1977 The Portland Forward Records Check of Crime Victims. Eugene: Oregon Research Institute.
Skogan, W.
1976 "Crime and crime rates." In W. Skogan (ed.), Sample Surveys of the Victims of Crime. Cambridge, MA: Ballinger.
1981 Issues in the Measurement of Victimization. Research report. Washington: U.S. Bureau of Justice Statistics.
Sparks, R., H. Genn, and D. Dodd
1977 Surveying Victims. New York: John Wiley and Sons.
Turner, A.
1972 The San Jose Methods Test of Known Crime Victims. Washington: National Criminal Justice Information and Statistical Service.
Yost, L., and R. Dodge
1970 "Household survey of victims of crime: second pretest—Baltimore, Maryland." Washington: U.S. Bureau of the Census, memorandum.

Explaining Information System Use with System-Monitored vs. Self-Reported Use Measures

JAMES S. ETTEMA

IT IS NOT yet clear whether the newly emerging information technologies will revolutionize communication. These technologies may revolutionize communication research, however, by generating data which is both different from and better than that previously available. In their extensive review of new data, Rice and Borgman (1983) point out that such new information systems as videotex can record such aspects of user behavior as the pattern of search strategies and search errors—aspects of behavior which the users themselves may not fully recognize. The authors also note that these systems can monitor such traditional communication variables as frequency and duration of information system use. These measures may well be more reliable and valid than self-reports of the behavior, though it is important to acknowledge that such measures are not without validity and reliability problems of their own.[1]

The availability of such new data offers communication researchers the

---

[1] In the research reported here, for example, most of the data from early in the field trial were destroyed or rendered suspect in the course of numerous hardware and software modifications. This report thus utilizes data from late in the field trial which the investigators in consultation with the system operator judged to be reliable and valid. Case et al.'s (1981) analysis of a similar information system provides another example of these sorts of problems.

Abstract New information technologies promise to generate new kinds of data. In this study, system-monitored use of a business-oriented videotex service is compared to self-reported use of the service. Results show that interest in the information provided by the service was the strongest correlate of the self-reports, while age and income were the strongest correlates of system-monitored use.

James S. Ettema is Associate Professor in the Department of Communication Studies, Northwestern University. This research was supported by Grant No. IST-8212164 from the Division of Information Science and Technology of the National Science Foundation.

Public Opinion Quarterly Vol. 49:381–387 © by the Trustees of Columbia University
Published by Elsevier Science Publishing Co., Inc. 0033-362X/85/0049-381/$2.50

opportunity not only to explore new aspects of behavior and reduce measurement error but also to understand more completely the nature of self-report data which remain very important to communication research. For example, "record check" studies can be conducted by comparing self-reports of communication behavior to system-monitored data which, for the purposes of the study, are assumed to be a more accurate record of that behavior. Such studies can estimate the degree to which self-reports converge (i.e., correlate) with "the record," and more important, can examine the consequences of using self-reports as compared to the more reliable and valid measures in analysis (Bernard et al, 1980; Katosh and Traugott 1981; Miller and Groves, 1983). Just such a study using self-reports and system-monitored usage data from a videotex field trial is reported here.

Research Setting, Method and Issues

The data presented here were generated during the study of a prototype videotex system which was tested on farm operators in three midwestern communities. During the eight-month field trial, the system allowed farmers to access a variety of information services including an agricultural market data service. This service provided regularly updated commodity price quotations from regional and national markets—just the type of information which potential system users had rated as the most important to their farming operations. This service was, then, the very model of the high-speed, specialized information retrieval system which many observers regard as the immediate future of the videotex information technology (Carey, 1982; Gollin, 1981; Sigel, 1983).

Self-reports on the use of this service were collected in a mail survey (N = 139, response rate = 74 percent) conducted near the end of the field trial. Farm operators were asked how often in the last month they and others in their household had used the service. Responses were recorded on a seven-point scale which ranged from "never" to "several times a day."[2] System-monitored use records made available by the system operator were aggregated into two system-monitored use variables: total time spent on the service for a period near the end of the field trial and total number of sign-ons for the same period.[3] Each of these variables thus provides a single measure of system use much the way a self-report is assumed to provide a "snapshot" of use for a particular period. The farm operators who chose to participate in the field trial were, of course, aware that such data would be collected.[4]

---

[2] Respondents were asked how often in the last month they (and others in the household) used the videotex system to "check the farm market prices." The response categories were: 1=never, 2=once a month, 3=several times a month, 4=once a week, 5=several times a week, 6=every day, 7=several times a day. In the following analysis, the "never" and "once a month" categories are collapsed as are the "every day" and "several times a day" categories to yield a five-point scale. The "last month" time frame of the question is longer than typical for media use self-reports but prior research on prototype information systems suggested that use of this system would be much lighter and more sporadic than use of traditional media. The time frame was adopted from Case et al. (1981).

The analysis of these measures assesses the consequences of employing self-reports as compared to system-monitored variables in attempting to explain use of this particular information system. Independent variables in this analysis include the standard demographics of age, family income, and education, which are typically found to correlate with informational media use. The independent variables also include two measures of user response to the market data service. One such measure is an index of user interest in the sort of information provided by the service. The other is an index of user satisfaction with the service.[5]

Analysis

The relationships between the independent variables and the three use measures are shown in Table 1. For the self-reported use measure, both age and education are significant, if modest, zero-order correlates. Thus, it was the younger and the better-educated farmers who tended to report heavier use of the service. However, it is clearly those farmers who expressed interest in market data who tended to report heavier use of the service. Multiple regression confirms the importance of interest—only the interest measure emerges from the analysis with a statistically significant beta.

For the "sign-on" system-monitored measure, the pattern of correlates is rather different. Age remains a significant zero-order correlate. However, family income replaces education as a significant correlate. Thus, it was the younger farmers who paid themselves less who tended to sign on to the system more often. Interest in the information offered is much less

---

[3] The system-monitored use records were collapsed into five-point categorical variables approximating the categorical self-reports. Means and standard deviations for all three use measures are shown in Table 1.

[4] See Rice and Borgman (1983) for a brief discussion of ethical issues raised by new data. Not the least of these issues is the impact of informed consent on user behavior. While the present study cannot address this issue directly, it is reasonable to assume that eight months into the field trial the farm operators had settled into patterns of system use which reflected their information needs, interests, and abilities rather than a concern for creating socially acceptable user records.

[5] The index of interest in the information provided by the service is composed of seven items which asked farm operators to rate the importance of various sorts of price data to their operation and the importance of the market data service in their decision to become users of the system. The index of user satisfaction is composed of four items which asked users to rate their satisfaction with the various sorts of price data provided by the service.

Table 1. Explaining Self-Reported and System-Monitored Information System Use

	Self-Reported Use		*System-Monitored Sign-ons*		*System-Monitored Time Spent*	
	(Mean=3.31, SD=1.29)		*(Mean=2.96, SD=1.44)*		*(Mean=2.95, SD=1.44)*	
	Zero-Order Correlation	*Beta (P Value)*	*Zero-Order Correlation*	*Beta (P Value)*	*Zero-Order Correlation*	*Beta (P Value)*
Demographic						
Age	.17	.06 (.48)	−.26	−.21 (.02)	−.23	−.16 (.07)
Income	.09	.05 (.59)	.26	−.24 (.01)	−.31	.28 (.00)
Education	.18	.08 (.37)	−.01	−.16 (.08)	−.04	−.11 (.19)
User response						
Interest	.42	.38 (.00)	.18	.15 (.07)	.21	.19 (.03)
Satisfaction	.10	.06 (.47)	.03	.04 (.68)	.09	.09 (.27)
		R^2 = .19		R^2 = .15		R^2 = .17

For this sample size (n = 131), correlations of r = .15 are statistically significant at the .05 level

strongly related to this system-monitored measure than to the self-report measure. The relative strength of the demographic variables as descriptors of system-monitored used is confirmed by the regression analysis—age and income both emerge with significant betas while interest has, at best, a marginally significant beta.

For the "time spent" system-monitored measure, the pattern of correlates is very similar to that of the sign-on measure. This, of course, was to be expected since these two measures are not independent—to spend time on the system it is necessary to first sign on. The measures are, in fact, highly correlated (r = .89). The regression analyses of these two measures differ just enough, however, to lead to slight differences in interpretational emphasis. For the "time spent" measure, income remains the best descriptor of use but age drops to marginal significance while interest rises to significance. Satisfaction with the system never emerges as significant correlate of any use measure.

To understand more about the nature of the self-report measure, it was reanalyzed with the system-monitored measures included as independent variables. These results, shown in Table 2, indicate that self-reports correlate with sign-ons and time spent at .29 and .34, respectively. This is about the same level as that reported by Case et al. (1981) in their analysis of a similar information system. Self-reports and system-monitored use are, then, related but they certainly do not measure exactly the same thing.

In the regression analysis, system-monitored use emerges as a significant descriptor of self-reports without much affecting the other variables in the equation. Thus, self-reports clearly do reflect system-monitored use even when a number of other variables are controlled but they even more clearly reflect interest in the information offered. Farmers who

Table 2. Explaining Self-Reported Information System Use with System-Monitored Use Measures as Independent Variables

	Zero-Order Correlation	*Beta (P Value)*	*Beta (P Value)*
Demographics			
Age	−.17	−.01 (.88)	−.02 (.88)
Income	−.09	−.01 (.90)	.03 (.71)
Education	.18	.11 (.18)	.11 (.20)
User response			
Interest	.42	.35 (.00)	.33 (.00)
Satisfaction	.10	.05 (.53)	.03 (.67)
System-monitored use			
Sign-ons	.29	.23 (.01)	–
Time spent	.34	– –	.27 (.00)
		$R^2 = .24$	$R^2 = .25$

were most interested in agricultural market information, it seems, tended to report that they used the system more than they really did.

Discussion

Somewhat different pictures of who used this information system emerge depending upon the particular measure of use. The picture emerging from the self-report measure shows interest in information to be the strongest correlate. This coincides with recent literature indicating that information interests are useful predictors of media effects (Ettema et al., 1983; Genova and Greenberg, 1979) and presumably media use as well. However, the picture which emerges from the system-monitored measures shows demographic variables to be the strongest correlates. The conclusion that the system was used by farmers "on the way up"—farmers who are younger and who currently pay themselves less, perhaps because they reinvest heavily in their farms—coincides reasonably well with data on the demographics of earlier adopters and heavier users of other new information technologies (Dozier and Rice, 1984).

Clearly, the choice of a use measure does have theoretical and methodological consequences. If the system-monitored measures are accepted as the more valid and reliable measures of use, then these results indicate that the self-reports tended to mask the impact of demographics. These results also suggest that information interests have some value as descriptors of "actual" use but that much of any correlation with self-reported use may be an artifact of method. Both self-reported information interests and self-reported use, after all, are drawn from the same survey questionnaire. The correlations may, then, reflect respondents' attempts to square their answers with each other or, more interestingly, the correlations may reflect a process of self-persuasion. That is, farmers may have persuaded themselves that if they really were interested in the information offered by the system, then they must have used it. Conversely, following Bem's (1965) "radical behaviorist" approach to attitude formation and change, farmers may have persuaded themselves that if they used the system then they must have been interested in it. It is, of course, not possible to decide between these alternatives using data from a single survey.

These results also indicate that satisfaction with the market data service was not a correlate of either self-reported or system-monitored use of the service. This finding is at odds with that of Case et al. (1981), who concluded that their self-reported use measures essentially *were* measures of satisfaction. Any attempt by respondents to square their answers with each other did not, then, extend to all parts of the questionnaire or, if the self-reports were somehow biased by interest, they were not similarly biased by satisfaction.

Overall, these results suggest that the new data made possible by the new technology will have important, if not revolutionary, consequences for our understanding of communication behavior. While the exact pattern of correlates of information system use seen here may or may not extend to other videotex services which serve other specialized audiences, such as securities brokers or upscale suburbanites, these results do suggest caution in the interpretation of relationships between supposed predictors of communication behaviors and self-reports of those behaviors. Caution is particularly imperative in the interpretation of relationships between self-reports of interest and self-reports of use. On the other hand, this and other recent research also suggests caution in the analysis of system-monitored data. Issues of ethics and accuracy are far from resolved. Further, such questions as what users *intended* to do and what they thought they were *doing* are interesting and important aspects of user behavior which no information system—other than the user's own brain—can monitor.

References

Bem, D. J.
1965 "An Experimental Analysis of Self-Persuasion." Journal of Experimental Social Psychology 1:199–218.
Bernard, H., P. Killworth and L. Sailer
1980 "Information Accuracy in Social Network Data IV." Social Networks 2:191–218.
Carey, J.
1982 "Videotex: The Past as Prologue." Journal of Communication 32:80–87.
Case, D., M. Chen, H. Daley, J. Kim, N. Mishra, W. Paisley, R. E. Rice, and E. M. Rogers
1981 "Stanford Evaluation of the Green Thumb Box Experimental Videotext Project for Agricultural Extension Information Delivery in Shelby and Todd Counties, Kentucky." Washington, D.C.: National Technical Information Service.

Dozier, D., and R. E. Rice
1984 "Rival Theories of Electronic Newsreading." In R. E. Rice et al. The New Media: Uses and Impacts. Beverly Hills, CA: Sage Publications.
Ettema, J. S., J. W. Brown and R. V. Luepker
1983 "Knowledge Gap Effects in a Health Information Campaign." Public Opinion Quarterly 47:516–527.
Genova, B. K. L. and B. S. Greenberg
1979 "Interest in News and the Knowledge Gap." Public Opinion Quarterly 43:79–91.
Gollin, A. E.
1981 "Consumers and Advertisers in the Electronic Marketplace: Implications for Newspapers and Other Media." Telecommunications Policy 5:171–180.
Katosh, J. P., and M. W. Traugott
1981 "The Consequences of Validated and Self-Reported Voting Measures." Public Opinion Quarterly 45:519–535.
Miller, P. V. and R. M. Groves
1983 "Measuring Reporting Accuracy of Criminal Victimization Through Record Checks." A paper presented to the American Association for Public Opinion Research Annual Meeting, Buck Hill Falls, PA.
Rice, R. E. and C. L. Borgman
1983 "The Use of Computer-Monitored Data in Information Science and Communication Research." Journal of the American Society for Information Science 34: 274–256.
Sigel, E.
1983 The Future of Videotext. White Plains, N.Y.: Knowledge Industry Publications.

Response Scales: Effects of Category Range on Reported Behavior and Comparative Judgments

NORBERT SCHWARZ, HANS-J. HIPPLER,
BRIGITTE DEUTSCH, AND FRITZ STRACK

IN SURVEY research as well as laboratory experiments category scales are frequently used to assess quasi-objective data such as frequency of church attendance, hours of television watched, and the like. Although a few practical guidelines are available (Payne, 1951; Sudman and Bradburn, 1982), the selection of the range of categories to be presented to respondents is usually left to the intuition of the researcher. Moreover, previous research has mostly concentrated on response biases without considering the cognitive processes underlying the use of category scales.

In the present paper we want to employ an information-processing perspective to explore the impact of the range of response categories provided to respondents on their behavioral reports and subsequent judgments. Central to the present perspective is the assumption that response scales are not simply "measurement devices" that respondents

Abstract Effects of the range of response categories provided in a closed answer format on behavioral reports and subsequent judgments were explored. Respondents reported their daily use of television along a scale that either ranged from "up to a half hour" "to more than two and a half hours" or ranged from "up to two and a half hours" to "more than four and a half hours." The former subjects reported less use of television than the latter and estimated the average use of TV to be lower. Moreover, the former subjects evaluated TV to be more important in their lives (Experiment 1) and reported less satisfaction with the variety of their leisure-time activities (Experiment 2). These results indicate that subjects inferred the average amount of television watching from the response alternatives provided them and used it as a standard of comparison in evaluating their behavior and its implications.

Norbert Schwarz is Assistant Professor of Psychology, and Brigitte Deutsch is Research Associate, Universität Heidelberg. Hans-J. Hippler is Project Director, Zentrum für Umfragen, Methoden und Analysen (ZUMA), and Fritz Strack is Assistant Professor of Psychology, Universität Mannheim. Experiment I was supported by ZUMA, Mannheim, as part of a larger study on response effects in surveys; Experiment II was supported by grant Schw 278/2-1 from the Deutsche Forschungsgemeinschaft to Norbert Schwarz and Fritz Strack. The authors want to thank George W. Bohrnstedt, Detlev Kommer, Seymour Sudman, and Robert S. Wyer, Jr. for their helpful comments on an earlier draft of this article. Address correspondence to Norbert Schwarz, Psychologisches Institut, Universität Heidelberg, Hauptstr. 47-51, D-6900 Heidelberg, Federal Republic of Germany.

Public Opinion Quarterly Vol. 49:388–395 © by the Trustees of Columbia University
Published by Elsevier Science Publishing Co., Inc. 0033-362X/85/0049-388/$2.50

use to report their behaviors. Rather, respondents may consider the range of behaviors described in the response alternatives to reflect the researcher's knowledge of or expectations about the distribution of these behaviors in the "real world." If so, respondents may use the range of behaviors described in the response alternatives as a frame of reference in estimating and evaluating their own behavior.

In selecting response categories, researchers are generally advised to "make sure that all reasonable alternative answers are included. Omitted alternatives and answers lumped into an "Other" category will be underreported" (Sudman and Bradburn, 1982: 21).

Moreover, Payne (1951) reported that "given a list of numbers (respondents) are prone to choose those near the middle of the list" (p. 80). He suggested that for knowledge questions that may involve guessing "it may be a wise precaution to set up any list of figures for a test of knowledge with the correct figure at one of the extremes" (p. 82). Indeed, most questions about behavioral frequencies are likely to involve guessing because records of the frequency with which respondents engage in "usual" behaviors are usually not available and respondents provide an estimate rather than an exact report. On the other hand, several studies (e.g., Carp, 1974; Payne, 1946, 1951; Quinn and Belson, 1969) indicated that respondents are particularly likely to endorse the first category of a list of three or more categories (verbal rating scales/evaluative items) provided them.

In the present studies we explore how respondents' reported estimate of the frequency with which they engage in a mundane behavior (watching TV) is affected by the range of the response categories provided in a closed answer format. In addition, we hypothesize that respondents will use the presumably "usual" behavior, inferred from the range of specific values stated in the response scales, as a standard of comparison in making judgments to which social comparisons are relevant (cf. Suls and Miller, 1977). Two experiments were designed to test these expectations.

Experiment I: Effects on Reported Behavior, Estimated Behavior of Others, and Importance of TV

In the first study, effects of category range on behavioral reports and estimates of the behavior of others were investigated. Respondents were asked to estimate the amount of television they watch along a category scale that ranged, in half-hour steps, either (a) from "up to a half hour" to "more than two and a half hours" or (b) from "up to two and a half hours" to "more than four and a half hours." It was hypothesized that respondents presented the high category range beginning with "up to two and a half hours" would report more TV consumption than respondents presented the low category range beginning with "up to a half hour." Previous research indicated that the average TV consumption in the Federal Republic of Germany is slightly more than two hours (Darschin and Frank, 1982). Thus, both versions of the response scale conform with Payne's (1951) suggestion to place the "true" response at the extreme of the list. In the present study, the "true" response was presented as the first value of the high category list and as the last value of the low category list. If respondents extract information about the real world from the range of the response categories provided, those presented the high category range list should estimate the average TV consumption to be higher as well. Moreover, they may use the inferred average behavior as a standard of comparison in evaluating how important TV is in their own leisure time.

METHOD

The first experiment was carried out in March/April 1983 as part of a larger survey (quota sample) with 132 German adults (more than 18 years of age). Quotas were based on an intersection of age, sex, and years of education. Respondents were randomly assigned to conditions and had to report how many hours they watched TV daily, using one of the two previously described scales, ranging in half-hours steps either from "up to a half hour" to "more than two and a half hours" or from "up to two and a half hours" to "more than four and a half hours." Subsequently, they estimated the daily TV consumption of the average citizen in an open answer format and evaluated how important a role TV played in their own leisure time on a scale from 0 (not at all important) to 10 (very important).

RESULTS AND DISCUSSION

An examination of respondents' behavioral reports indicates that those who were presented the low range scale tended to choose categories in the middle of the list, whereas respondents who were presented the high range scale tended to endorse the first category provided (see Table 1). To assess the impact of the response scale on subjects' reports, their responses in both conditions were coded as either (a) two and a half hours or less or (b) more than two and a half hours. The range of the response scale affected respondents' behavioral reports as expected. Specifically, only 16.2 percent of the respondents who were presented the low category range reported watching TV for more than two and a half hours, while 37.5 percent of the respondents presented the high category range did so, chi^2 (1) = 7.7, $p < .005$. Moreover, respondents presented the low category range subsequently estimated the average TV consumption of German citizens to be 2.7 hours, whereas respondents presented the

Table 1. Reported Behavior, Estimated Behavior of Others, and Importance of TV as a Function of Category Range

Reported Behavior			
Low Category Range		*High Category Range*	
Up to 1/2h	7.4% (5)	Up to 2 1/2h	62.5% (40)
1/2h to 1h	17.7% (12)	2 1/2h to 3h	23.4% (15)
1h to 1 1/2h	26.5% (18)	3h to 3 1/2h	7.8% (5)
1 1/2h to 2h	14.7% (10)	3 1/2h to 4h	4.7% (3)
2h to 2 1/2h	17.7% (12)	4h to 4 1/2h	1.6% (1)
More than 2 1/2h	16.2% (11)	More than 4 1/2h	0.0% (0)
	100% (68)		100% (64)
Estimated TV use of average citizen	2.7h		3.2h
Importance of TV	4.6		3.8

NOTE: N = 132. Importance of TV was assessed on a 10-point bipolar scale (1 = not at all, 10 = very important).

high category range estimated it to be 3.2 hours, $F(1,127) = 8.8$, $p < .004$, indicating that they based these estimates in part on the frequencies suggested by the range of the scales. Consistent with the comparison information derived from these scales, respondents presented the low category range (suggesting to them that other people watch little TV) evaluated TV to be *more* important in their *own* leisure time, $M = 4.6$, than respondents presented the high category range, $M = 3.8$, $F(1,127) = 7.92$, $p < .006$. Note that this was true even though the former reported watching *less* TV than the latter.

In summary, the data provide strong support for the assumption that the range of categories presented as part of a closed format question would affect respondents' behavioral reports as well as related judgments. However, in the present study respondents were explicitly asked to estimate average TV consumption. This task might have directed their attention to the category range for appropriate hints, resulting in a high salience of the category range. As a result, the effect of category range on the evaluation of TV importance might be more pronounced than it would have been otherwise. This possibility was explored in a second study.

Experiment II: Effects on Leisure Time Satisfaction

Two procedures were used in this study to guard against artifactual effects of making the response range salient. First, respondents estimated the average TV consumption after rather than before giving evaluative judgments. Second, for half of the respondents the category scale and the evaluative judgment were separated by filler items to render the category range less salient. Finally, a control condition was run in which subjects were not presented a category scale but reported TV consumption in response to an open ended question. In summary, the present study followed a 3 × 2 factorial design with scale (low or high category range, open question) and filler items (present, absent) as between-subjects factors.

Respondents' satisfaction with the variety of their leisure-time activities (rather than importance of TV) was assessed as an evaluative judgment in this study. Since watching a lot of TV is socially undesirable in West Germany (Mueller, 1980), it was hypothesized that respondents would report higher satisfaction with the variety of their leisure-time activities when the response scale suggests that other people watch more TV than when it suggests others watch less TV.

METHOD

In a survey embedded in a larger questionnaire, 79 German employees, recruited during office hours and randomly assigned to conditions, reported their daily TV consumption on one of the two previously described category scales, or in response to an open question. Subsequently, they reported their satisfaction with the variety of their leisure-time activities on a scale from 1 (very dissatisfied, I wish there were more variety) to 11 (very satisfied, I don't want more variety). For some respondents the two questions were separated by several filler items regarding their mass media consumption while for others they were not. Finally, respondents estimated the TV consumption of the average German citizen.

RESULTS AND DISCUSSION

As in Experiment I, the ranges of the response scales affected subjects' reports of their own use of TV. Using two and a half hours as a cut-point, 19.2 percent of the control group respondents reported watching TV for more than two and a half hours per day while none (0 percent) of the respondents presented the low category range and 29.6 percent of those presented the high category range reported doing so, chi^2 (2) = 8.68, $p < .02$. As in the previous study, respondents presented the low category range preferred the middle categories while those presented the high category range preferred the first category presented, as shown in Table 2.

Moreover, the type of scale used by subjects to report their behavior affected their evaluation of their leisure-time activities, $F(2,73) = 3.13$, $p < .05$. Specifically, planned comparisons indicated that respondents who reported their TV consumption on the high range category scale,

Table 2. **Reported Behavior, Leisure Time Satisfaction, and Estimated Behavior of Others as a Function of Category Range**

Reported Behavior					
Low Category Range		*High Category Range*		*Open Answer Format*	
Up to 1/2h	11.5% (3)	Up to 2 1/2h	70.4% (19)	Up to 1/2h	0.0% (0)
1/2h to 1h	26.9% (7)	2 1/2h to 3h	22.2% (6)	1/2h to 1h	19.2% (5)
1h to 1 1/2h	26.9% (7)	3h to 3 1/2h	7.4% (2)	1h to 1 1/2h	15.4% (4)
1 1/2h to 2h	26.9% (7)	3 1/2h to 4h	0.0% (0)	1 1/2h to 2h	46.2% (12)
2h to 2 1/2h	7.7% (2)	4h to 4 1/2h	0.0% (0)	2h to 2 1/2h	0.0% (0)
More than 2 1/2h	0.0% (0)	More than 4 1/2h	0.0% (0)	2 1/2h to 3h	19.2% (5)
	100% (26)		100% (27)		100% (26)
Estimated TV use of average citizen	2.8h		3.7h		3.7h
Leisure time satisfaction	8.2		9.6		8.8

NOTE N = 79. Leisure-time satisfaction was assessed on an 11-point bipolar scale (1 = very dissatisfied, I wished there were more variety, 11 = very satisfied, I don't want more variety).

suggesting to them a high average TV consumption, subsequently reported more satisfaction with the variety they had in their leisure time, $M = 9.6$, than respondents presented the low category range, $M = 8.2$, $t(73) = 2.50$, $p < .02$. Control group subjects responding in an open answer format reported intermediate satisfaction, $M = 8.8$, that did not significantly differ from either of the experimental groups, t's $(73) = 1.34$ and 1.15, respectively, *n.s.* Finally, these effects were not affected by the presence of filler items, $F < 1$ for main effect and interaction.

As expected, the type of scale presented also affected respondents' estimates of the TV consumption of the average German citizen, $F(2,73) = 6.17$, $p < .003$. Specifically, respondents presented the low category range estimated the average consumption to be lower, $M = 2.8$, than both control group respondents and respondents presented the high category range, M's $= 3.7$, t's $(73) = 3.0$ and 3.12, respectively, $p < .004$. The latter groups did not differ from one another, suggesting that the information transmitted by the high category range was similar to control group respondents' spontaneous assumptions concerning the typical use of TV. Again, the results were not affected by the presence of filler items, $F < 1$ for the main effect and $F(2,73) = 1.11$, *n.s.* for the interaction.

In summary, the data replicated the results of the first study and supported the hypothesis that respondents derived comparison information from the category range they were presented and used this comparison information in making satisfaction judgments. Moreover, the effects on satisfaction judgments were obtained without any attempt to direct respondents' attention to the presumed typical behavior; that is, their estimates of the "typical" behavior were not assessed until after satisfaction judgments had been reported. Finally, neither the effects of response range on satisfaction judgments nor its effects on estimates of others' behavior were affected by the introduction of filler items, suggesting that the effects are rather robust.

General Discussion

The present studies demonstrate that response scales serve informative functions. Specifically, the response categories provided in a closed answer format inform the respondent about the researcher's knowledge of or expectations about the real world. The response categories suggest a range of "usual" or "expected" behaviors, and this information affects respondents' behavioral reports as well as related judgments. Specifically, respondents seem to use the frame of reference provided by the scale in estimating their own behavior and they seem reluctant to report behaviors that are unusual in the context of the response scale, e.g., because they constitute the extreme categories.

In addition, respondents use the presumably "usual" behavior suggested by the response scale as a standard of comparison in related judgments. In line with this assumption, respondents' evaluations of the importance of TV in their own lives as well as of the variety of their leisure-time activities were affected by the comparison standards suggested by the scales. Because the impact of the information provided by the scale should be the more pronounced the less other relevant comparison information is available to respondents, effects of category range are likely to be obtained in content domains for which direct access to comparison information is difficult. The effects should be much smaller, however, for content domains in which the behavior of others is highly visible. This implication remains to be tested in future research.

Turning to the applied implication of the results, researchers should consider the use of open answer formats in obtaining data on behavioral frequencies. As Sudman and Bradburn (1982: 115) have noted, "There is no difficulty in coding such responses, since the data are numerical and can easily be processed without need for additional coding." Precoding the question, on the other hand, may introduce systematic bias in respondents' behavioral reports and related judgments, because response scales are not only measurement devices but serve informative functions as well.

References

Carp, F.M.
1974 "Position effects on interview responses." Journal of Gerontology 29:581–587.
Darschin, W., and B. Frank
1982 "Tendenzen im Zuschauerverhalten. Teleskopie-Ergebnisse zur Fernsehnutzung im Jahre 1981." Media Perspektiven 4:276–284.
Mueller, Albrecht
1980 "Einstellungen der Fernsehzuschauer zur weiteren Entwicklung des Mediums Fernsehen." Media Perspektiven 3:179–186.
Payne, Stanley L.
1946 "Some opinion research principles developed through studies of social medicine." Public Opinion Quarterly 10:93–98.
1951 The Art of Asking Questions. Princeton: Princeton University Press.
Quinn, S.B., and W.A. Belson
1969 The Effects of Reversing the Order of Presentation of Verbal Rating Scales in Survey Interviews. London: Survey Research Center.
Sudman, Seymour, and Norman M. Bradburn
1982 Asking Questions. San Francisco: Jossey-Bass.
Suls, J.M., and R.L. Miller (eds.)
1977 Social Comparison Processes: Theoretical and Empirical Perspectives. New York: Hemisphere Publishing.

The Polls: America's Most Important Problems Part II: Regional, Community, and Personal

TOM W. SMITH

In Memory of George Gallup, Sr.

THE most important problem question helps us understand both the broad strokes and fine details of recent American history. It illuminates important secular trends such as the emergence and disappearance of civil rights as a social concern and gauges the impact on the public consciousness of specific events from riots to recessions. The parallel series on sectional, community, and personal problems add even greater depth to our historical and sociological understanding, especially by demonstrating that problems frequently differ across frames of reference. The most important problem questions are invaluable for both portraying and explaining recent American history and the social change American society has been undergoing.[1]

Tracking trends on the sectional, community, and personal problem questions (Tables 1, 2, and 3) add considerable information about the post-World War II period (each of these series starts in the mid to late forties about a decade after the national most important problem series). These variants are not merely national concerns writ small. Typically, these variants bring to mind quite different sets of problems. Rather than having only one salient problem and expressing that in national, regional, local, and personal contexts, people often chose different problems as the reference changes.[2] The largest and most obvious

[1] For some analyses that have been conducted with these items see the literature cited in Douglas A. Hibbs, Jr., "The mass public and macroeconomic inflation," American Journal of Political Science 23: 705–31; Tom W. Smith, "America's most important problem: a trend analysis, 1946–1976," Public Opinion Quarterly 44: 164–180; Michael Bruce MacKuen, "Social communication and the mass policy agenda," in Michael Bruce MacKuen and Steven Lane Coombs (eds.), More Than News: Power in Public Affairs, Beverly Hills: Sage Publications, 1981.

[2] A detailed comparison was made between national and non-national variations that were either asked on the same survey (19 comparisons) or on surveys separated by a month or less (7). This included comparison with national most important problem questions from Roper, Harris, NORC, and SRC-Michigan that are not included in the tables.

Tom W. Smith is Senior Study Director, National Opinion Research Center (NORC), University of Chicago.

America's Most Important Problem, Part I: National and International, appeared in the Summer 1985 issue of *POQ*.

Public Opinion Quarterly Vol. 49:403–410 © by the Trustees of Columbia University
Published by Elsevier Science Publishing Co., Inc. 0033-362X/85/0049-403/$2.50

difference is that foreign affairs are infrequently mentioned on the non-national versions. Only once are foreign affairs chosen by as many as 10 percent on any of these questions and usually are mentioned by less than a tenth as many as on the national question. Even when we exclude foreign problems (and Don't Knows, which are more common on non-national questions), there are still major differences among domestic concerns. The sectional, local, and community questions all collect considerably more mentions in the miscellaneous categories. These consist of such matters as transportation, education, the environment, and housing that are rarely chosen on the national question. The personal problem question also varies by having considerably more mentions of health and such vague categories as "personal" and "family" matters.

The comparatively short time series for sectional problems (1949–1964) shows two major trends. Economics is the main concern in the early 1950s, but declines as worries over civil rights increase. In addition, housing steadily decreases from being the number two problem in 1949 until it practically disappears by 1964. On the community question the disappearing category is transportation. It ranks first in 1949, but rattles around third place from 1961 to 1975. The other major trend on community problems is the drop of economics from first place during the 1961–1968 period to second behind social control after 1968. In stark contrast to the fluctuation in the top concern on the national, sectional, and local question, economics has always been the most important personal problem. Despite the civil rights movement, the social disorder of the sixties, and the energy crises of the seventies, people always placed their pocketbook first when they consider their personal problems.

The data were compiled with the assistance of the Roper Center, University of Connecticut; the Louis Harris Data Center, University of North Carolina; and the National Opinion Research Center, University of Chicago. These archives supplied data from the following survey organizations:

AIPO - American Institute of Public Opinion (Gallup Poll)
ELEC - Election Studies, Survey Research Center, University of Michigan
HARRIS - Louis Harris and Associates
NORC - National Opinion Research Center, University of Chicago
POS - Public Opinion Survey (Gallup)
ROPER - Roper Organization

Most Important Sectional Problem

Question Wording	*Surveys*
A. What do you think is the most important problem facing this section of the country today?	AIPO436, 443K, 444, 564, 689, 695, 696
B. What do you think is the most important problem facing this section of the country—that is, the [name of geographic region] today?	AIPO454, 518, 528

Table 1. Most Important Sectional Problem

Survey	AIPO436	AIPO443K	AIPO444	AIPO454	AIPO518
Date	(1/49)	(6/49)	(6/49)	(3/50)	(7/53)
Foreign affairs	.030	.037	.025	.040	.081
Domestic	.847	.850	.858	.864	.747
Economic	.413	.511	.507	.474	.374
Social control	.027	.011	.011	.036	.016
Civil rights	.018	.052	.056	.028	.051
Government	.012	.011	.017	.018	—
Housing	.168	.114	.103	.087	.065
Misc., named	.137	.123	.159	.181	.180
Unspecified	.072	.028	.005	.040	.061
Not classified	.019	.018	.012	—	—
Don't know	.105	.094	.103	.087	.173
None	—	—	—	.010	—
N cases	3184	1327	2825	1458	1532
N responses	3366	1704	2988	1535	1639
Responses/cases	1.06	1.28	1.06	1.05	1.07

Survey	AIPO528	AIPO564	AIPO689	AIPO695	AIPO696
Date	(3/54)	(4/56)	(4/64)	(7/64)	(8/64)
Foreign affairs	.026	—	.010	.007	.014
Domestic	.836	.845	.832	.821	.823
Economic	.521	.341	.296	.241	.245
Social control	.037	.038	.026	.035	.039
Civil rights	.047	.134	.275	.316	.284
Government	—	.003	—	—	—
Housing	.023	.039	.013	.009	.010
Misc., named	.192	.195	.129	.111	.157
Unspecified	.016	.095	.093	.109	.088
Not classified	.010	.008	—	.006	.007
Don't know	.113	.131	.157	.167	.156
None	.015	.016	—	—	—
N cases	1562	1934	3509	3515	3513
N responses	1671	2094	3844	3754	3942
Responses/cases	1.07	1.08	1.10	1.07	1.12

Most Important Community Problem

Question Wording	*Surveys*
A. What would you say is this city's worst problem today?	GALLUP Special Survey, 1948
B. What do you think are the most important problems here in this community?	AIPO641
C. What do you think is the most important problem facing this immediate vicinity, that is, this neighborhood today?	AIPO697
D. What do you think is the most important problem facing this community today?	AIPO720, 725
E. What is the most important problem of this community?	NORC4018
F. In your opinion, what is the most important problem facing this community today?	AIPO757
G. What would you say are the two or three major problems facing communities like your own today?	HARRIS2224, 2355
H. What do you regard as your community's (your city's) worst problem?	AIPO861, 931

Table 2. Most Important Community Problem

Survey	Gallup Special Survey	AIPO641	AIPO697	AIPO720	AIPO725
Date	(5/48)	(2/61)	(8/64)	(11/65)	(3/66)
Foreign affairs	—	—	.010	—	.062
Domestic	.910	.780	.734	.723	.734
Economic	.126	.431	.198	.191	.209
Social control	.112	.032	.029	.088	.085
Civil rights	.011	.048	.133	.098	.082
Government	.062	.017	—	.031	.017
Transportation	204	.042	.067	.034	.052
Housing	.146	.036	.022	.044	.041
Ecology	.085	.016	.018	.064	.059
Energy	—	—	—	—	—
Misc., named	.105	.118	.136	.124	.114
Unspecified	.059	.040	.131	.049	.075
Don't know	.091	.220	.255	.277	.205
None	—	—	—	—	—
N cases	D.K.	2873	4003	3521	3555
N responses	D.K.	3300	4325	3782	4045
Responses/cases	D.K.	1.15	1.08	1.07	1.14

Survey	NORC4018	AIPO757	HARRIS2224	AIPO861	HARRIS2355	AIPO931
Date	(3-6/67)	(1/68)	(8/72)	(12/72)	(1/74)	(6/75)
Foreign affairs	.001	.069	.003	—	—	—
Domestic	.899	.803	.961	.857	.933	.876
Economic	.170	.199	.156	.121	.189	.187
Social control	.116	.196	.402	.255	.281	.248
Civil rights	.043	.062	.034	.036	.021	.025
Government	.022	.029	.036	.058	.033	.062

Transportation	.101	.072	.046	.095	.082	.078
Housing	.069	.036	.039	.029	.070	.033
Ecology	.083	.040	.078	.082	.029	.065
Energy	—	—	—	—	.098	—
Misc., named	.271	.169	.170	.164	.130	.148
Unspecified	.024	—	—	.017	—	.030
Don't know	.076	.081	.024	.084	.038	.063
None	.024	.049	.012	.058	.031	.060
N cases	3095	1500	3218	1504	1594	1558
N responses	5009	1705	5982	1736	2708	1813
Responses/cases	1.62	1.14	1.86	1.15	1.70	1.16

Most Important Family Problems

Question Wording[3]	*Surveys*
A. What is the most important problem that you and your family face today?	AIPO362, 378, 387T, 392, 473 516
B. What do you think will be the biggest problem *you or your family* will have to face in the next few years?	NORC153
C. What would you say is your biggest worry these days—the thing that disturbs you the most?	AIPO478
D. Everybody has problems which worry them and their families. What would you say are the two or three biggest problems that face you and your family today?	ROPER74
E. What would you say is the one thing you are worrying about most these days?	AIPO488
F. What would you say is your greatest worry these days—the thing that disturbs you the most?	AIPO535
G. What is the most urgent problem facing you, or your family, today?	AIPO548
H. Now I'd like to ask some questions about your personal cares and concerns these days. As far as you are personally concerned, what are the two or three main things that have concerned or worried you and your family in the past few months? Anything else?	HARRIS1301
I. When you think of your own situation and that of your family, what problem concerns you most?	NORC4018
J. What is the most urgent problem facing you and your family today?	AIPO752

---

[3] All versions refer to family, daily lives, and/or personal problems except C, E, and F. These versions tend to fall between national problems and family problems in their wording and responses.

Question Wording	*Surveys*
M. And what would you say are the two or three most serious personal problems people like yourself must deal with today?	HARRIS2224, 2355
N. We like to have people tell us, in their own words, what sort of problems they have to deal with in their daily lives. Can you tell me what some of the problems are that you face these days in your own life? Of all you've told me, what would you say is the single most important (personal) problem you face?	ELEC72
O. What is the most important problem facing your family at this time?	AIPO980
P. We like to have people tell us, what sorts of problems they have to deal with in their daily lives. Can you tell me what some of the problems are that you face these days in your own life? Anything else? Of all you've told me, what would you say is the single most important problem you face?	ELEC74, 76 [omits "own" and has no follow-up question on "single most important] problem"
Q. What do you consider to be the biggest problem in your own personal life?	AIPO1005

Table 3. Most Important Family Problems

Survey	AIPO362	AIPO378	AIPO387T	AIPO392K	NORC153
Date	(12/45)	(9/46)	(1/47)	(3/47)	(10/47)
Foreign affairs	.024	.011	—	—	.044
Domestic	.863	.921	.884	.922	.860
Economic	.599	.659	.611	.608	.603
Social control	.002	—	—	—	—
Civil rights	.004	—	—	—	—
Government	—	—	—	—	—
Housing	.138	.119	.132	.133	.115
Old age	.001	.020	.017	.022	—
Health	.055	—	—	—	—
Family/personal	.026	.084	.085	.126	.124
Misc., named	.002	.028	—	—	—
Unspecified	.036	.011	.039	.033	.018

Don't know	.027	.024	.027	.024	.097[a]
None	.087	.045	.088	.055	—
N cases	2909	3232	1466	1437	1287
N responses	3030	3506	1504	1504	1454
Responses/cases	1.04	1.08	1.03	1.05	1.13

Survey	AIPO473	AIPO478	ROPER74	AIPO488[b]	AIPO516	AIPO535	AIPO548
Date	(3/51)	(8/51)	(1/52)	(3/52)	(5/53)	(8/54)	(6/55)
Foreign affairs	.080	.216	.115	.116	.052	.065	—
Domestic	.901	.724	.693	.739	.820	.811	.734
Economic	.673	.466	.522	.404	.651	.390	.455
Social control	—	.012	—	—	.002	—	—
Civil rights	—	.001	—	—	.003	—	.004
Government	—	—	—	.042[c]	—	—	—
Housing	.045	—	.044	—	.035	—	.041
Old age	.002	—	.015	—	.015	—	.007
Health	.049	.039	.042	.098	.058	.126	.103
Family/personal	.040	.117	.039	.121	.032	.176	—
Misc., named	.008	.080	.031	.023	.024	.015	.098
Unspecified	.084	.009	.092	.051	—	.104	.026
Don't know	.018	.059	.082	.046	.039	.028	.135
None	—	—	.019	.099	.089	.097	.131
N cases	2101	2057	2495	2072	1549	1626	1462
N responses	2149	2158	3395	2172	1609	1765	1486
Responses/cases	1.02	1.05	1.36	1.05	1.04	1.09	1.02

Survey	HARRIS1301	NORC4018	AIPO752	HARRIS1813	AIPO791	HARRIS2224
Date	(3/64)	(2-7/67)	(10/67)	(3/68)	(10/69)	(8/72)
Foreign affairs	.011	.036	.093	.081	.063	.005
Domestic	.808	.896	.745	.710	.781	.907
Economic	.289	.339	.520	.473	.596	.375
Social control	.009	.030	.021	.074	.022	.116
Civil rights	.031	.007	.025	.047	—	—
Government	—	—	—	.010	—	—
Housing	.029	.035	.021	—	—	.023
Old age	.052	.023	.024	—	—	.030
Health	.187	.189	.062	—	.049	.049
Family/personal	.039	.126	—	—	—	.141
Misc., named	.095	.147	.046	.106	.047	.092
Unspecified	.077	—	.026	—	.067	.081
Don't know	.020	.005	.014	.209	.067	.025
None	.162	.061	.146	—	.091	.061
N cases	1419	3074	1585	1586	1543	3218
N responses	1940	5104	1770	2065	1623	4863
Responses/cases	1.37	1.66	1.12	1.30	1.05	1.51

[a] This also includes "None."

[b] Respondents who replied "not at all" to the filter question, "In general, how much would you say that you worry—a lot, a fair amount, a little, or not at all?" were not asked the most important problem follow-up. They are coded "none" in Table 3. If they are excluded the distribution would be foreign affairs—.128, economic—.449, government—.047, miscellaneous, named—.026, unspecified—.057, and don't know—.051 ($N = 1856$, responses = 1956).

[c] "Government" here includes problems that would be coded elsewhere if separately distinguished (e.g., debt of the country, condition of the country).

Survey	ELEC72	HARRIS2355	ELEC74	ELEC76	AIPO980	AIPO1005
Date	(11/72)	(1/74)	(10/74)	(11/76)	(7/77)	(6/78)
Foreign affairs	.004	—	.006	—	—	—
Domestic	.887	.925	.911	.874	.781	.776
Economic	.402	.407	.530	.542	.558	.350
Social control	.033	.074	.024	.021	—	—
Civil rights	.011	.014	.008	—	—	—
Government	—	.022	—	.003	—	—
Housing	—	.018	—	.002	—	—
Old age	—	.033	—	—	.014	.019
Health	.091	.056	.083	.083	.063	.126
Family/personal	.221	.107	.177	.148	.042	.194
Misc., named	.117	.104	.088	.050	.024	—
Unspecified	.012	.090	.001	.025	.080	.087
Don't know	.001	.027	—	.003	.036	.039
None	.109	.049	.083	.124	.183	.184
N cases	1050	1594	2523	2410	1516	1505
N responses	1623	2645	4060	3316	1656	N.D.
Responses/cases	1.55	1.66	1.61	1.38	1.09	1.03